Real-time feedback control of gene expression

Jannis Uhlendorf

Real-time feedback control of gene expression

A DISSERTATION PRESENTED
BY

JANNIS UHLENDORF

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN THE SUBJECT OF
SYSTEMS BIOLOGY

UNIVERSITÉ PARIS-DIDEROT (PARIS 7)
FONTIÈRES DU VIVANT DOCTORAL SCHOOL (ED 474)
PARIS, FRANCE
APRIL 2013

DATE OF THE DEFENCE: APRIL 19th 2013

MEMBERS OF THE JURY:

PRESIDENT	CATHERINE DARGEMONT
RAPPORTEURS	JOHN LYGEROS
	DIEGO DI BERNARDO
EXAMINERS	PASCAL SILBERZAN
	GUY-BART STAN
ADVISORS	GREGORY BATT
	PASCAL HERSEN

Real-time feedback control of gene expression

Abstract

Gene expression is fundamental for the functioning of cellular processes and is tightly regulated. Inducible promoters allow one to perturb gene expression by changing the expression level of a protein from its physiological level. This is a common tool to decipher the functioning of biological processes: the expression level of a gene is changed and one observes how the perturbed cell behaves differently from an unperturbed cell. A shortcoming of inducible promoters is the difficulty to apply precise and time-varying perturbations. This is due to two reasons: (i) cell-to-cell variability and noise in the gene expression process which limit the precision of the applied perturbations. (ii) the difficulty to quantitatively predict the behavior of cellular systems on a long time-horizon, which would be required to apply time-varying perturbations. But precise time-varying perturbations are particularly informative about the dynamics of a biological system.

Here I present a feedback control platform, that can control the expression of a gene in yeast cells with quantitative accuracy over long time periods. The platform integrates fluorescence microscopy to monitor gene expression, microfluidics to act on the cellular environment and software implementing real-time image analysis and feedback control. This closed-loop control setup is able to drive the expression of a yeast gene in a population of cells or in a single cell for both time-constant and time-varying target profiles. I use the high osmolarity glycerol (HOG) pathway, a stress response pathway of *Saccharomyces cerevisiae*, to activate gene expression and I show that the platform can be modified to use other gene induction systems. In addition to the gene expression control platform, I present a feedback control system able to control the activity of the HOG pathway.

Understanding cellular dynamics in a quantitative way necessitates the ability of applying precise perturbations. The gene expression control platform that I present here is a major step towards this goal, since it allows to precisely perturb the expression level of a protein.

Contrôle temps réel en boucle fermée de l'expression génétique

Résumé

L'expression génétique est un processus cellulaire fondamental réglé de manière fine. Les promoteurs inducibles permettent de perturber l'expression génétique en changeant l'expression d'une protéine par rapport à son niveau physiologique de référence. Cette propriété en fait un outil incontournable pour décrypter le fonctionnement des processus biologiques via la comparaison du comportement de la cellule sous divers niveaux d'induction. Toutefois, une limite actuelle à l'utilisation des promoteurs inducibles provient de la difficulté à appliquer des perturbations précises et dynamiques. Les deux obstacles principaux étant: (i) la variabilité intercellulaire ainsi qu'à la nature aléatoire de l'expression génétique qui limite la précision de la perturbation appliquée. (ii) la difficulté à prèdire quantitativement le comporteement des systèmes biologiques sur les longues periodes requises pour des objectifs d'expression variables dans le temps. Or des perturbations précises et changeant dans le temps permettent d'obtenir de riches informations sur la dynamique d'un système biologique.

Est présenté ici une plate-forme de contrôle temps réel en boucle fermée qui permet le contrôle quantitatif sur une longue durée de l'expression génétique chez la levure. Cette plate-forme utilise la microscopie par fluorescence pour suivre l'expression génétique, un système microfluidique pour interagir avec l'environnement cellulaire ainsi qu'un logiciel permettant l'analyse d'image en temps réel et le calcul de la stratégie de contrôle à appliquer. Ce système permet le contrôle de l'expression d'un gène chez la levure, tant au niveau d'un population cellulaire qu'au niveau de la cellule seule et ceci pour un objectif d'expression constant ou dépendant du temps. Le système de réponse au chocs hyper-osmotiques de la levure *S. cerevisiae* (HOG pathway) a été utilisé pour influencer l'expression génétique. Toutefois, la possibilité d'utiliser un autre système d'induction sans profondes modifications de la plate-forme est démontrée. De surcroît au développement de cette plate-forme est également ici démontré la possibilité de contrôler le système HOG.

Afin de comprendre la dynamique cellulaire et de pouvoir la quantifier, il est nécessaire de pouvoir appliquer des perturbations précises. La plate-forme de contrôle de l'expression génétique présentée ici permet de perturber avec précision le niveau d'expression d'une protéine et représente donc une contribution majeure dans cette direction.

Contents

Acknowledgments

First I would like to thank the members of my PhD committee, Catherine Dargemont, John Lygeros, Diego di Bernardo, Pascal Silberzan and Guy-Bart Stan. I am profoundly grateful that you invested your precious time in reading this manuscript, traveling to Paris and evaluating my defense.

The two persons I cannot thank enough at this place are my two supervisors Gregory Batt and Pascal Hersen. Thank you for your outstanding support during the last $3\frac{1}{2}$ years. Thank you for all the ideas, the discussions, explanations and for providing a delightful working atmosphere. Thank you also for always being there when I needed help and for motivating me in difficult times. It was really a pleasure working with you.

A special thanks goes to Agnès Miermont who taught me experimental work when I arrived to Paris and had barely touched a pipette before. I also owe a big thank you to Xavier Dupportet for helping me with the DNA constructions. Thank you Clément Vulin for the many ideas and for the nice time in the lab.

Thank you Stéphane Douady and Sébastien Leon for taking the burden of being in my Thesis Advisory Committee and for all the tips you gave me. Thank you also for your always rapid and competent help with yeast problems, Sébastien.

I would like to thank Samuel Bottani for the countless discussion we had about this project and for the guidance he provided.

Thank you, Gilles Charvin for inviting me to your laboratory and for sharing your microfluidic device with me. A special thanks goes to Thierry Delaveau, who took the time to show me how to transform yeast cells. Thank you François Fages for supporting me during this work and for all the great suggestions you made. I would like to express my great appreciation to Eugenio Cinquemani

for the various pleasant discussions and explanations. Also I would like to acknowledge the help provided by Romain Bourdais and his internship students Alexandre Peyre, Thomas Michel and Xiaohong Wu.

Aishah Prastowo I would like to thank for taking the risk of being my first internship student, for the contribution she made to this work and for the fruitful discussion we had. Joé Schaul I would like to thank for his great work on modeling the osmotic stress response and for all the discussions. Thank you, Artémis Llamosi for helping me out with the French abstract. I would also like to thank Blair Wilfred Lebert for his help with the English language.

In addition I would like to thank Antoine Decrulle, Szymon Stoma, Frédéric Devaux, Mathilde Garcia, François Bertaux, Stephanie Aubin, Carole Barache, Nadine Beyer, Lucie Bouchu, Danielle Champeau and Nadia Mesrar and everybody else that I forgot.

Last but not least I would like to thank all the members of the CONTRAINTES team as well as of the group of Pascal Hersen for the pleasant working atmosphere.

1 Introduction

1.1 Motivations

Proteins are the building blocks of a cell. They define the physical shape of a cell, catalyze reactions and mediate intra- and intercellular communication; in short, they are fundamental for cell functioning. Protein production is controlled by gene expression, which is regulated at various levels. This allows the cell to adjust its protein production to the current needs despite constant perturbations. Both internal factors, such as the cell cycle stage or growth rate, and external factors, such as the physico-chemical environment, may influence gene expression. Regulating gene expression is of vital importance for the cell: for example, a cell would not be able to accommodate to changes in its environment without changing the expression level of certain genes. Regulating gene expression is also a requirement for multicellular life, because it allows differentiation so that cells having the same genotype can still express different sets of proteins.

1.1.1 Gene expression is regulated

This central role of gene expression explains the effort that has been undertaken within the last 50 years to understand the underlying mechanisms. The first mechanism of gene expression regulation was discovered by Jacques Monod and François Jacob in 1961 [1], when they unraveled the mechanism by which the lac operon is regulated in *E. coli* [2]. The lac operon contains three enzymes that are required for the metabolism of lactose. The expression of these enzymes is controlled by a protein called the lac repressor, which binds to the deoxyribonucleic acid (DNA) in the absence of lactose and hinders the expression. The repressor dissociates from the DNA in the presence of lactose, meaning that the enzymes are only produced when they are needed. Since the discovery of the lac operon, many other mechanisms of gene regulation have been found at every level of gene expression [3] and are still being found. For example, Wan *et al.* recently discovered that in *S. cerevisiae* RNA transcripts which are degraded in response to heat shock often change their conformation in response to a temperature change from 30°C to 37°C [4]. This result suggests that there might be a previously unknown regulation mechanism, in which RNA transcripts act as thermo-sensors.

Feedback loops are an abundant feature of biological systems

Regulations can either have a positive or negative influence. In a positive feedback loop, the regulation has an indirect positive influence on its own activity, while negative feedback loops show the opposite behavior. An example of a negative feedback loop is the lac operon: an increase in the level of lacZ, the enzyme which metabolizes lactose, leads to a decrease in lactose concentration, which in turn leads to a down-regulation of this enzyme (see Figure 1). Negative feedback loops are a very common feature in biology [6] as they ensure a balance between the available amount of the substrate of an enzyme and the concentration of this enzyme. In addition negative autoregulations can stabilize the expression level of a protein. Becskei *et al.* showed in a synthetic system, that a gene which inhibits its own expression shows a lower variation compared to the unregulated gene [7]. Cells use such negative autoregulations to maintain the expression level of proteins in the face of external perturbations [8]. In addition to a negative feedback loop, the lac operon also contains a positive feedback: activation of the operon induces the membrane lactose transporter lacY, which in turn positively influences the amount of lactose which is imported [9]

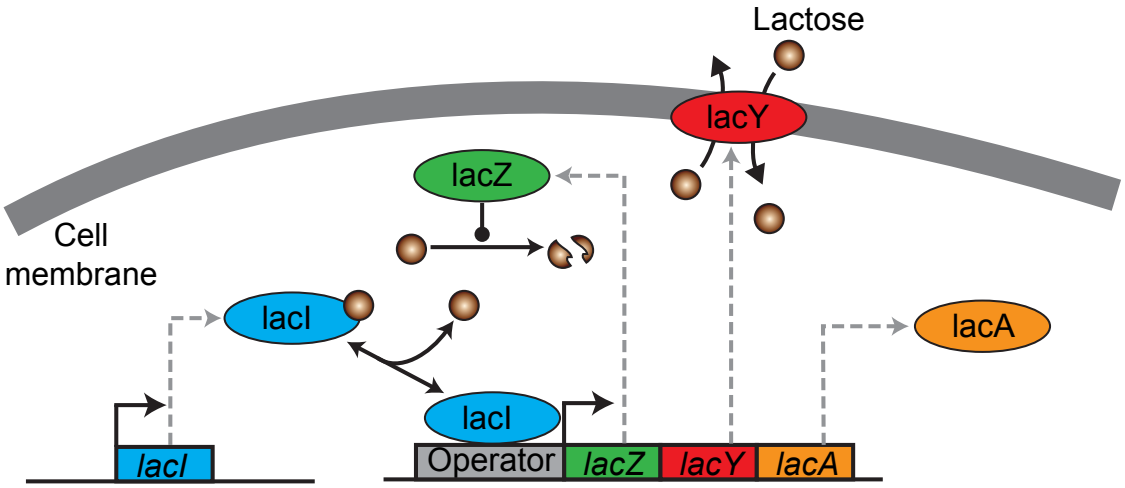


Figure 1: Regulation of the lac operon. The lac operon consists of the three genes *lacZ*, *lacY* and *lacA* which code for proteins involved in the import and metabolism of lactose. *lacZ* codes for the enzyme β-galactosidase which breaks down lactose to galactose and glucose, while *lacY* codes for a membrane transporter responsible for the uptake of lactose (lactose permease). The third gene, *lacA*, codes for the enzyme thiogalactoside acetyltransferase whose biological function is still unclear [5]. The transcription factor lacI (lactose repressor) binds to the lac operator in the absence of lactose thereby preventing expression of the operon. This creates a negative feedback loop, because expression of the lac operon has a negative influence on the concentration of lactose.

and thereby as well the expression of the operon. Positive feedback loops can lead to bistability, meaning that for a certain input the system can be found in one of two possible states. For the lactose system this means that for low lactose concentrations two states are possible: (i) The expression level of the operon is close to zero because not enough transporter is available to import enough lactose such that the system could be activated. (ii) The operon is expressed, because there is already enough transporter to reach a critical threshold of internal lactose concentration. Whether the system is in state (i) or (ii) depends to the history of the system. Such a switch like behavior ensures that the enzymes for lactose consumption are only expressed once the lactose concentration reaches a certain threshold.

1.1.2 Gene expression can be controlled externally

The ability to alter gene expression plays an important role in the pharmaceutical and biotechnical industries. For example, insulin can nowadays be produced by *E. coli* cells, which carry the human insulin gene [10]. Before it was possible to produce genetically engineered human insulin, the method of producing insulin analogs was to isolate them from the pancreas of pigs or cows, however this animal-sourced insulin has a slightly different amino acid sequence than human insulin.

The expression level of a gene is mainly determined by its promoter sequence, a DNA sequence which is the initial binding site for the RNA polymerase and transcription factors, which influence the expression of the gene. A promoter not only influences the strength of the expression of a gene but also its temporal expression profile and potential co-regulations with other genes. To adjust the enzyme level, the promoter expressing the insulin gene had to be carefully chosen. To provide researchers as well as genetic engineers with the possibility to choose a promoter suitable for their specific requirements, promoter libraries have been developed for different organisms [11–14].

An important tool in biology is the ability to influence the expression of certain genes externally. This allows the study of the biological role of the influenced gene by playing with its expression level (*e.g.* dependence of the growth rate on the concentration of a certain protein [15]). Inducible promoters allow one to influence the expression of a gene via external factors. They are often small molecules that interact with transcription factors or are sensed via membrane receptors, which trigger a signaling cascade that affects gene expression. In addition there are promoters which react to physical changes like temperature shifts [16, 17] or light [18]. Note that in fact all these promoters are not influenced directly by small molecules, temperature, or light, their activity change relies on transcription factors whose binding affinity to the promoter is affected in response to a small molecule, temperature shift or, light intensity change.

Inducible promoters can be endogenous promoters which are already present in the studied cell, however the natural promoter is used to control a different gene. This can be a problem since the environmental factor controlling the promoter is often not limited to a particular gene, but controls a number of genes or has other side effects, an effect called pleiotropy. To circumvent side effects one can introduce gene inducible systems which originate from a distinct species, so that the effect of the inducer is limited to the gene to be controlled (orthogonality) [19]. Limitations of inducible promoters are pleiotropic effects mentioned above as well as the stochastic nature of

gene expression, which renders it hard to predict the effect of an inducer at the single cell level [20, 21].

1.1.3 Synthetic feedback within the cell

Synthetic biology is a novel field of biological research that aims at implementing new biological functions into living organisms. From a scientific point of view, this approach offers unprecedented opportunities to put our knowledge about biological processes to a test, because one could argue that we only really understand a system if we are able to reconstruct it. From an engineering perspective, synthetic biology promises great economic impact in the near future, because microorganisms could be used for many applications like the production of bio-fuels, pharmaceutical developments and drug production, or as biological sensors.

The first synthetic gene networks constructed were a genetic toogle switch [22] and a network exhibiting oscillations consisting of 3 genes [23]. Since then, many other synthetic gene networks have been developed, including improved oscillators, logic gates or communication modules.

The exact influence of an inducible promoter is often not easy to predict. One way to tackle this lack of precision of inducible gene systems is to implement a synthetic feedback control in which the expressed protein inhibits its own expression. There have been several attempts to implement such a feedback system in live cells. In 2000 Becskei and Serrano implemented a feedback loop in *E. coli* in which a protein inhibits its own expression [7] (see Figure 2A). They observed a much lower variation for the autoregulated system as compared to a non-regulated version of the system. A negative feedback system in mammalian cells was implemented in 2011 within the group of Diego di Bernardo [25]. This systems contains two negative feedback loops, one acting via the transcriptional repressor tetR-KRAB[1] and another via a short-hairpin RNA[2] (see Figure 2C). The strength of both feedback loops can be tuned externally and the authors propose this setup to implement a toogle switch, which shows a bistable behavior. Also in 2011, Stapleton *et al.* implemented a feedback control of protein expression in mammalian cells which acts at the translational level [24]. A modified ribosomal L7Ae protein binds its own mRNA thereby inhibiting its own translation as well as the translation of a potential fusion protein. The mRNA is expressed

[1] TetR-KRAB is a fusion protein of the tetracycline repressor (tetR) and of the the Krüppel-associated box repressor (KRAB) protein

[2] Short hairpin (sh)RNAs are RNA sequences that can bind an mRNA and silence its translation

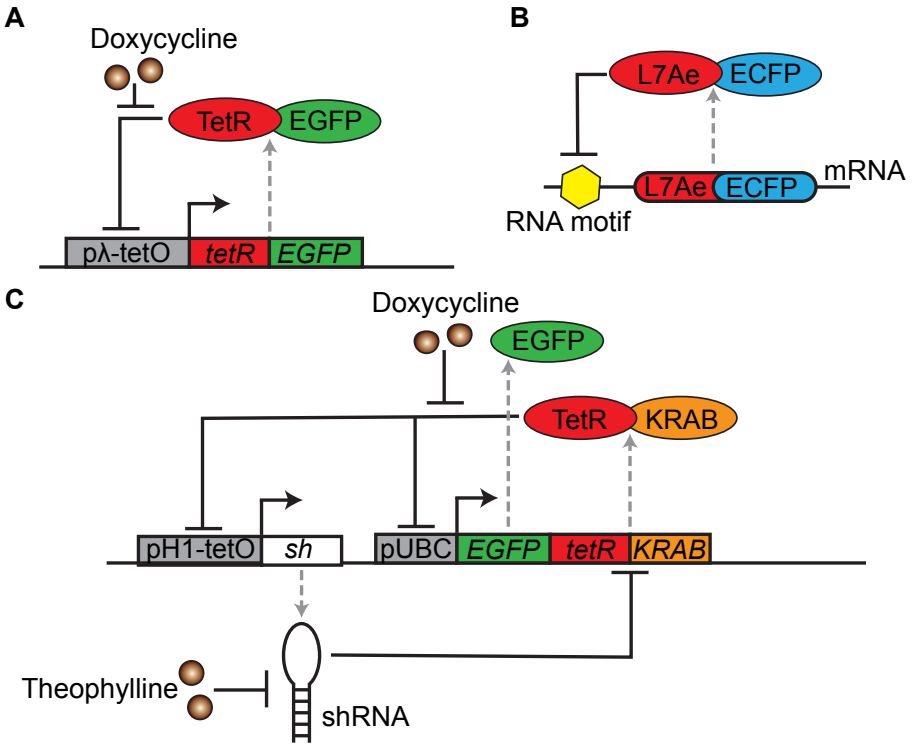


Figure 2: Synthetic biology feedback loops. **(A)** Synthetic feedback loop by Becskei *et al.* [7] in which the expression of a fusion of the tetracycline repressor (tetR) and of a green fluorescent protein (EGFP) inhibits its own expression. TetR binds to the tet operator (tetO) sequence within a modified λ-promoter in the absence of Doxycycline. If Doxycycline is present, this inhibition is relieved. **(B)** Translation-based feedback loop implemented by Stapleton *et al.* [24]. An mRNA carries a fusion protein of the ribosomal binding protein L7Ae and of a cyan fluorescent protein (ECFP). In addition the mRNA carries a sequence to which L7Ae binds, thereby inhibiting its own translation as well as the translation of the ECFP fusion protein. **(C)** Feedback system by Polynikis *et al.* [25] in which the fusion protein tetR-KRAB inhibits the expression of a short hairpin (sh)RNA, of an EGFP protein and of itself. The tetR protein binds the tetO sequence which is part of the H1 promoter and the KRAB protein acts as a transcriptional repressor acting on a genomic region of 2-3 kilobases, thus inhibiting the expression of the shRNA and of the EGFP/tetR/KRAB mRNA. EGFP and the tetR-KRAB fusion protein are driven by the constitutive ubiquitin promtoter (pUBC), but because the location of this promoter is close to the H1 promoter, it is affected by the binding of tetR-KRAB to H1. This binding can be prevented by the addition of doxycycline. The shRNA was designed to bind the mRNA sequence of tetR-KRAB, thereby inhibitxing translation of tetR-KRAB. The binding strength of the shRNA can be modulated by the addition of theophylline.

using a constitutive promoter, therefore no external control of the expression is possible. In this approach the strength of the feedback is adjustable by changing either the sequence of L7Ae or of the K-turn RNA motif to which L7Ae binds (see Figure 2B).

1.1.4 Reverse engineering of biological systems

Quantitative biology is a field of biological research that aims at describing biological processes in a quantitative manner, rather than in the classical qualitative manner. The advancement of this field has been driven by the improvements of methods to quantitatively observe cellular processes (*e.g.* fluorescent proteins) as well as by the increased usage of mathematical modeling to describe cellular dynamics. The lack of precision in conventional gene induction systems is a problem for quantitative biology, because understanding the quantitative dynamics of cellular processes requires the ability to apply precise perturbations to the system [26]. A classical way of system identification is to perturb a system and to monitor the systems response to this perturbation. For example, biological systems are often perturbed by genetic knockouts, thereby removing one regulator from the system. Slightly more dynamic perturbations can be applied using inducible promoters, which allow one to express genes conditionally and, to some extent, to gradually change the expression level of a gene. Even more informative are time-varying perturbations, because they allow one to explore the dynamic behavior of a system [27]. In principle, time varying perturbations could be applied by inducible promoter systems, but for several reasons this would not be effective in practice. First, it is hard to predict the precise action of an inducer due to cell to cell variability and the inherent stochasticity of gene expression. In addition, such an approach would require a detailed model of the dynamics of the induced protein in order to predict the actual protein level at different times. Without monitoring the expression level of the perturbed protein, significant deviations between the actual and the desired protein profile are to be expected. However, it becomes feasible to impose precise, time-varying perturbations on the level of a protein by constantly observing the level of the perturbed protein and by adjusting the level of induction based on this observation. In the following I will present such a feedback control system of gene expression, which uses an external feedback loop to circumvent the problem of predicting the precise effect of an inducer.

1.1.5 External synthetic feedback

An alternative approach to the implementation of a feedback system inside of the cell is to work with an external feedback loop. In a feedback control system (closed loop controller), the controlled process is observed, such that deviations between the control objective and the observed behavior of the system can be corrected by modifying the control strategy. Automated control is heavily used in technical systems. A simple example is a refrigerator which constantly measures the temperature inside its cooling chamber and adjusts the cooling power based on this observation and the settings provided by the user. Thus the temperature in the cooling chamber is kept constant even if the external temperature varies. In control engineering we distinguish between open loop control, in which the controller does not observe the system, and closed loop (or feedback) control, where the controller is provided with some real time measurement of the system (see Figure 3). Open loop control requires a good model of the controlled system and of its environment, because in this control framework errors in the control strategy cannot be compensated based on real-time observations. Because precise quantitative models of gene expression systems are rare and gene expression shows a high stochasticity [28–30], closed loop control, in which the controller updates the control strategy based on observations, is the method of choice to control gene expression with a satisfactory precision. Advantages of utilizing an external feedback loop instead of an internal one to control gene expression are versatility and robustness with respect to varying environmental conditions or unmodeled system dynamics. An external control loop makes it possible to easily change the control target and in addition allows for dynamic targets, that vary over time. As well, the control setup can be modified by adjusting the control software, whereas for a synthetic feedback genomic modification are necessary to alter the control system.

A feedback control system for gene expression

The requirements for the implementation of an external closed loop control of gene expression are (i) the ability to observe gene expression and (ii) the ability to act on gene expression. A method to observe the amount of a certain protein inside a cell is the use of fluorescent proteins [31], which have the advantage that the measurement can be done on living cells and at the single cell level. The simplest method to influence gene expression is the use of inducible promoters which react to small compounds that are added to the media of the cells. To be able to dynam-

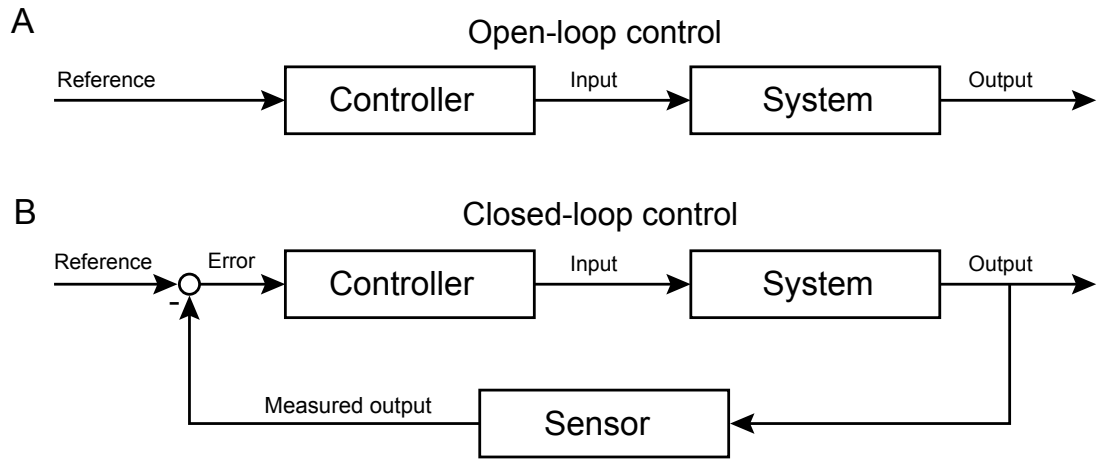


Figure 3: Open-loop and closed-loop control. **(A)** An open-loop controller computes the input to the system based only on the given reference and on the estimated state of the system. No feedback is used to compensate for control errors. **(B)** A closed-loop controller measures an output of the system and uses this measurement to compute the input, thereby dynamically correcting for control errors.

ically influence gene expression, a controller needs to add and remove the inducible compound from the cellular media. This is possible *a priori* with microfluidic devices in which cells can grow but are constrained in their movement while fluid media is flowing through the device, providing the cells with nutrients. By changing the composition of this flowing media, substances can be both added to or removed from the cellular environment, while all compounds produced by the cells are washed away, keeping experimental conditions constant.

The topic of this thesis is the development of such a closed loop control system for gene expression in the yeast *Saccharomyces cerevisiae*. As a system to induce gene expression I will use the high osmolarity glycerol (HOG) pathway (see Figure 4), which is a stress response pathway in yeast, activated by an increase in osmolarity of the cellular media, which allows cells to adapt to changes in the osmolarity [32]. An increase of the osmolarity leads to a volume loss of the cell due to water leaving the cell and the HOG pathway restores the cell volume by promoting the synthesis of glycerol, thereby increasing the cellular osmolarity. Activation of the HOG pathway mediates the upregulation of almost 600 genes [33]. One of the genes most strongly and specifically activated by osmotic shock is *STL1* [34, 35], which codes for a glycerol proton sym-

porter[3] in the membrane. In this work, I will use the promoter of *STL1* to drive the expression of a yellow fluorescent protein. Hyperosmotic stress has a quite severe effect on the cell, because the expression of many genes is changed. In addition the pathway incorporates several natural feedback mechanisms which deactivate the pathway after prolonged activation. Using a complex signaling pathway which includes feedback mechanisms to control gene expression is not an easy task, but it was chosen in this context to demonstrate that control is feasible even in a complex system, where many phenomena cannot be modeled or influenced externally.

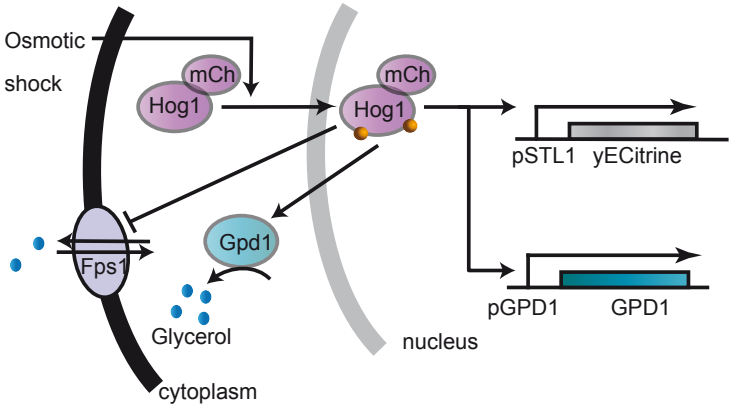


Figure 4: Schematic representation of the HOG pathway, the system I will use to activate gene expression. An osmotic shock triggers the phosphorylation of the protein Hog1 which in response translocates to the nucleus where it activates the expression of various genes. One of the activated genes is *STL1* whose coding sequence has been replaced here by the fluorescent marker yECitrine. The pathway incorporates natural feedback mechanisms, which are mediated mainly via the increased production of glycerol. One of the mechanisms involved is the enhanced expression of the glycerol producing enzyme GPD1. In addition the membrane glycerol transporter Fps1 closes in response to osmotic stress.

1.2 Contributions

At the time when I started this thesis (in 2009), no real-time control of gene expression had been demonstrated. Developing a real-time control platform is not an easy task, because of different

[3]A symporter is a type of membrane transporter which transports two different molecules in the same direction. This allows one to utilize a concentration gradient of one molecule (often an ion) to drive the transport of the other molecule.

technical obstacles that have to be addressed: the real-time control of a cellular process using fluorescent proteins as a means to observe the controlled process requires automated and robust image analysis. For traditional microscopy experiments, image analysis can be done after the experiment in a semi-automated fashion, whereas for closed loop control the images have to be analyzed in real-time. For the control of single cells, each cell has to be identified automatically and tracked over time. It is very important that these analyses work in a robust manner, because the controller works without manual intervention. In addition image analysis, tracking, and the computation of the control strategy have to be performed within the time in which two consecutive images are taken, which limits the choice of methods that can be applied here. Last but not least the control software needs to be able to drive the microscope as well as to control the microfluidic device. I implemented a closed loop control platform in MATLAB[4], which can drive a fluorescent microscope via the open source software Micromanager [36], perform automated image analysis, cell-tracking and computation of the control strategy in real time and is able to control the microfluidic device. Since the control software could prove useful for other closed loop control systems in biology, I made it available online[5]. In particular the cell-tracking algorithm I developed could be useful for other projects which rely on tracking yeast cells in time-lapse movies.

A closed loop control system for gene expression is a useful tool to probe the dynamics of cellular processes, because it allows one to apply precise and time varying perturbations to the expression level of a gene (see Figure 5). In this case, the protein level which is the output of the controlled system would act as an input to the investigated system. This could for example be applied to study the functioning of gene networks by applying oscillatory inputs [37]. Here the temporal expression of a transcription factor could be controlled and it could be observed how the system reacts to inputs of different frequencies. Similar approaches have led to new insights in the dynamics of signaling cascades [38, 39], but for gene networks such analyses have not been feasible until now due to the lack of methods to impose time varying profiles on input signals. Another use of such a system is to study noise in gene expression, by fixing the expression level of the controlled gene to a certain value and observing the variability. This would for example allow one to determine the gene expression noise depending on the gene expression level. In addition,

[4]The MathWorks, Natick MA, USA

[5]http://www.msc.univ-paris-diderot.fr/~jannis/control_code

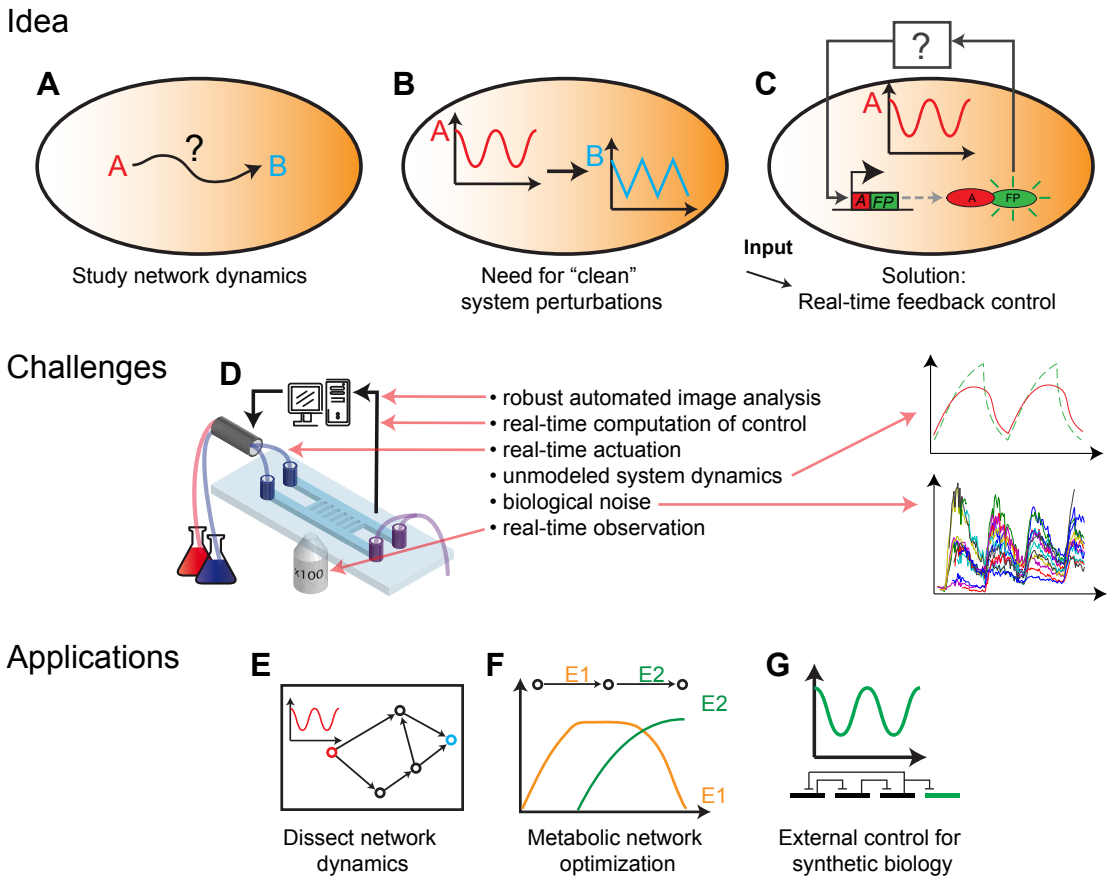


Figure 5: Motivations for controlling gene expression. **(A-B)** Often we are interested in understanding the effect that a protein A has on the dynamics of another protein B. One way to study the influence that protein A has on B is to apply precise time-varying perturbations to the concentration of A and to observe the response of protein B. **(C)** Precise time-varying perturbations of a protein level can be achieved using real-time feedback control. Protein A is fused to a fluorescent protein (FP) and controlled by an inducible promoter. A feedback controller (black box) decides, based on the measurement of the A-FP fusion protein, which inputs to apply such that the concentration of protein A will follow a desired profile. **(D)** Real-time feedback control is challenging for a biological system. **(E-G)** Applications of a feedback controller for gene expression: **(E)** The dynamics of biological networks could be investigated by controlling the concentration of one network node and observing the response of the other network nodes. **(F)** The yield of metabolic networks could be optimized by ensuring a specific temporal expression profile for certain enzymes. **(G)** Features that are implemented today into cells in the filed of synthetic biology (*e.g.* stable oscillations) could be tested or even replaced by external feedback control.

feedback control of biological processes could have implications in biotechnical industries. For example in metabolic engineering, where cells are being optimized in their production of a certain substance, the timing of the expression of an enzyme is often crucial. In addition, over-expression of the enzyme might have toxic effects. In such a situation controlling the expression of one or multiple protein(s) in the cell could help to optimize the metabolic production process.

The closed loop control platform I present in this thesis is an important step towards these goals. In addition it is an ideal tool to perform real-time experiments on yeast cells, real-time meaning that the experimental conditions can depend on a readout of a cellular behavior. This could for example be used to synchronize the onset of a stimulus with cellular events like cell division or the upregulation of a certain protein.

In 2011 we showed that it is possible to control the signaling activity of the HOG signaling cascade *in vivo* using a simple proportional-integral (PI) controller [40], which was the first real-time control approach for a signaling pathway. Activation of the HOG pathway leads to the phosphorylation of the protein Hog1 which is subsequently imported into the nucleus where it influences the expression of various genes [32]. Using a fluorescent label for Hog1, we constantly monitored the Hog1 nuclear localization and controlled it using a feedback controller. Because of the natural feedback mechanisms within the HOG pathway only a limited set of control targets could be reached.

Later in 2011 two other papers demonstrating feedback control of a cellular process were published. Toettcher *et al.* used optogenetics to control the activity and localization of a signal transduction protein (PI3K[6]) in eucaryotic cells at the single cell level [41]. Also using optogenetics, Milias-Argeitis *et al.* were able to control the expression of a gene in a population of yeast cells within a chemostat [42] to a constant target value. The two control approaches as well as a comparison to my work will be discussed in Section 7.4.

In 2012 we presented a feedback control framework for gene expression [43], which is able to control the expression of a yeast gene in both populations and single cells. This control platform allows the control of gene expression in both populations of cells or single cells with either time-constant or time-varying target functions at high precision. The range between the lowest and highest feasible target value spans one order of magnitude. Because the time-scale on which gene expression changes is much longer than the time-scales for signal transduction pathways,

[6]Phosphoinositide 3-kinase

the platform has been designed for long-time experiments (>15 hours), while being able to automatically and robustly detect and track cells over this time. The platform is based on a model predictive controller (MPC), which uses a model of the gene expression system to predict the behavior of the system for different inputs.

1.3 Outline

This document is organized as follows. In Chapter 2, I will present background information about the HOG pathway as well as an overview about methods to induce gene expression in yeast. Chapter 2 also contains a brief introduction in feedback control systems and state estimation. In Chapter 3 Materials and Methods, I present methods to detect and follow cells in movies and give an overview about the microfluidic devices used in this work. The development of a feedback control system for the HOG signaling cascade and a first computational investigation of how the HOG system can be used to control gene expression are presented in chapter 4. Chapter 5 describes the gene expression control platform and provides the main results of this thesis. In Chapter 6, I show how this platform can be extended to use another method of gene induction which is based on the methionine inducible promoter. A discussion of this work as well as a comparison to related work and conclusions can be found in Chapter 7.

2

Background

In this thesis I describe the development of a platform to control gene expression in yeast using the HOG pathway as a mean to influence gene expression. Because this is an essential part of my work, I give a short introduction on yeast cells, their osmotic stress response and means to introduce gene expression here. In addition I will introduce basic notions of feedback control, focusing on the two control methods I will apply later, proportional integral derivative (PID) control and model predictive control (MPC). This chapter is not intended to give an extensive summary of these different fields, but instead aims at providing a brief introduction for readers not familiar with these subjects.

2.1 Yeast and its response to osmotic stress

2.1.1 Yeast as a model organism

Yeasts are unicellular eukaryotic microorganisms that belong to the class of Fungi. Yeasts are among the oldest domesticated species, since they have been used for thousands of years for baking and brewing. In the 1830s it was first recognized that yeast cells are living organisms and Theodor Schwann who was studying alcoholic fermentation in yeast at that time called them Zuckerpilz (German for sugar fungus), because he observed the ability of beer yeast to convert sugar to alcohol [44, 45]. In 1838 Julius Meyen translated this name to Latin which gave rise to the name *Saccharomyces cerevisiae* [45] (*saccharum* is Latin for sugar, *myce* means fungus and *cerevisiae* means of beer). Schwann was not the first one to investigate alcoholic fermentation. In 1789 Antoine Lavoisier was the first person giving an estimate of chemical changes occurring in alcoholic fermentation [46]. Between 1855 and 1875 Louis Pasteur finally deciphered the role of yeast in alcoholic fermentation and was the first one to distinguish between aerobic and anaerobic utilization of sugar [46]. Today *S. cerevisiae* is one of the most important model organisms in biology. Reasons for this are its fast reproduction cycle ($\sim$ 90 minutes) and the ease of genetic manipulation. In addition *S. cerevisiae* is non-pathogenic and simple to grow in a laboratory environment. In 1996 it was the first eukaryotic organism for which the whole genome had been sequenced [47]. The yeast genome can be easily altered by homologous recombination, which allows the introduction of foreign DNA sequences at specific locations in the genome. This makes it possible to delete genes by replacing them with a marker gene (see Section 2.1.2) or to tag proteins with fluorescent markers. Consequently, a gene deletion library including most of the approximately 6000 yeast genes [48, 49], as well as a library tagging each protein with a fluorescent marker [50, 51] have been constructed. The biological role of about 85 % of the protein coding genes in *S. cerevisiae* have attributed functions [52] and the Saccharomyces Genome Database[1] (SGD) provides manually curated information for most genes.

2.1.2 Genetic manipulation in yeast

One reason for the success of *S. cerevisiae* as a model organism is the ease of genomic manipulation in this organism. Synthetic DNA can easily be introduced using different methods like

[1] http://www.yeastgenome.org

electroporation or lithium acetate treatment. In addition *S. cerevisiae* shows a high frequency of integration of external DNA into the genome, in particular if the introduced DNA is linear [53].

Novel genes can be introduced in *S. cerevisiae* either by integrating them into the genome or they can rest on plasmids within the cells. The process of introducing exogenous genetic material into yeast cells is called transformation. In both methods a selectable marker is required to select for cells which carry the novel gene. This marker is usually a resistance to a certain drug or a gene that complements a specific auxotrophy[2]

Expression of a gene from a plasmid bears two problems. First the plasmid can be lost during division and second the copy number of the plasmid in the cell is neither known nor constant over time. While the first problem can be overcome by using plasmids carrying a centromere (CEN plasmids), which ensures that each daughter cell maintains at least one plasmid during division, the second problem can only be solved by integrating the DNA sequence into the genome.

Integration of DNA sequences into the genome is possible via homologous recombination, a cellular process responsible for the repair of DNA double-strand breaks and for crossover during meiosis. Homologous recombination is a genetic exchange between pairs of homologous DNA sequences. It can be used to introduce exogenous DNA fragments at specific locations in the genome. A sequence homology to the genomic location to be altered is placed in the DNA to be introduced. Positive clones can then be identified with a selective marker.

Cellular signaling

The natural environmental conditions of unicellular organisms like yeasts are subject to permanent fluctuations. Factors like temperature, pH or nutrient levels and sources vary with time in a natural environment and the cell needs to cope with these changes [54–56]. An example is the transfer of yeast cells from a media containing glucose as the main energy source to a media containing galactose. To metabolize galactose, the cell has to convert it first to glucose-1-phosphate, which then can be used for energy production in the glycolysis, one of the main metabolic pathways. The enzymes which catalyze this pathway are only expressed if galactose is present and glucose is not present [57]. This can be understood in the light of evolution. Cells need to spend their resources economically in a competitive environment, so they produce enzymes only when

[2]Auxotrophy is a term for the inability of an organism to synthesize a particular compound which is vital for the cell.

they actually need them.

Cells also need to respond to signals which do not pass the cellular membrane. This is achieved by signaling pathways that are activated by cell surface receptors, which sense external signals and in turn activate one ore more intracellular signaling pathways. In addition to the cell surface receptors there are cytoplasmic receptors, which respond to molecules that can cross the cellular membrane. A signaling pathway transmits the sensed information to intracellular effector proteins which mediate the appropriate response. Signaling pathways are often comprised of a chain of signaling proteins which activate each other in cascade. The function of signal transduction is to amplify signals, to integrate different stimuli and of course to transfer information to the appropriate location inside the cell.

2.1.3 Osmotic stress response in yeast

In the following section I will describe the osmotic stress response in yeast, which I will use as a means to activate gene expression in my control setup. A major reason for choosing this pathway in favor of other gene activation systems was that it incorporates several feedback mechanisms which render it difficult to control the pathway, allowing me to demonstrate that feedback control of biological systems is even functional in difficult cases where not all system dynamics can be accurately modeled.

Osmotic shock

An example of an environmental factor to which cells need to react is a change of the osmolarity of the surrounding media. The osmolarity of a solution depends on the number of molecules solved in a certain volume of the solution. The measure of osmolarity is osmoles per liter (osmol/L), which is defined as the number of osmotically active particles (in moles) per liter within a solution. An osmotic pressure arises if two solutions with different osmolarities are separated by a semipermeable membrane, that is a membrane which is permeable only to certain molecules. For example the cell membrane of yeast forms such a semipermeable membrane: water can diffuse freely through the membrane, while most small and large molecules cannot. If at some point the osmolarity inside the cell is higher than the one of the external media, water will flow inside the cell such that the osmotic pressures inside and outside the cell are balanced. This increases the osmotic pressure. On the other hand, if the osmolarity is higher in the external media, the cell will

loose water (see Figure 6). Cell volume regulation is vital for cells. In case the turgor [3] pressure rises too high (hypoosmotic shock), the cell is in danger of bursting. On the other hand, a drop in cell volume (hyperosmotic shock) can also affect the functioning. First of all the drop in volume affects the concentration of all molecules within the cell. In addition it has been shown that in response to severe osmotic shocks, the concentration of solubles within the cell can reach so high that their diffusion is affected due to crowding [58]. Both hyper- and hyposomotic shocks have severe impacts on cellular functioning and the cell needs to deal with these in order to be viable. Two distinct signaling pathways mediate the cellular response to hyper- and hypoosmotic shock.

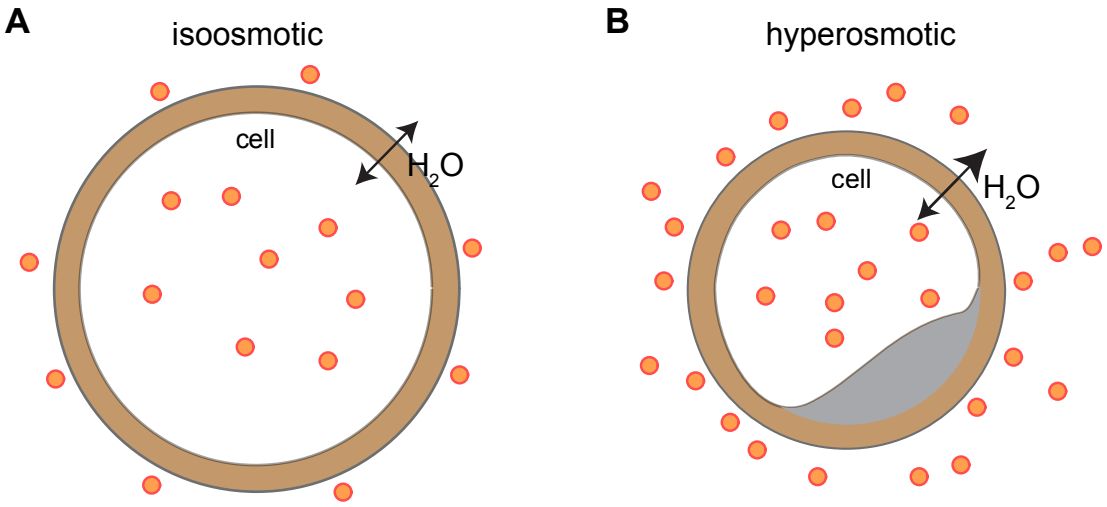


Figure 6: Cell size change after an osmotic shock. **(A)** Under isosmotic conditions the intracellular osmolarity is only slightly higher than in the external media, which creates a hydrostatic pressure (turgor pressure) which pushes outward on the cell wall (brown). The orange dots represent osmotically active compounds. **(B)** In hyperosmotic conditions, the external osmolarity is significantly higher than the normal one, which leads to an efflux of water from the cell (the turgor pressure decreases). In condition of severe stress, the cytoplasm can even partly detach from the cell wall (plasmolysis). Sound functioning of cellular processes requires to restore the osmotic balance and thereby the cell volume to its original size.

[3] Turgor pressure describes a pressure that acts from inside on the cell membrane and is caused by the osmotic flow of water inside the cell which is driven by the high solute concentration inside the cell.

Adaptation to osmotic stress: The HOG pathway

As mentioned above, the main problem of an elevated osmolarity in the media for the cell is a loss of water and the resulting shrinkage of the cell volume. The adaptation to the hyperosmotic stress is mainly mediated by the accumulation of the small osmolyte glycerol, thereby increasing the internal osmolarity and restoring the water balance. Production of glycerol is mediated by the high osmolarity glycerol (HOG) pathway. I will give a quite detailed description of the HOG pathway here, because its dynamics are of importance for the control approaches I will present in this thesis. In Chapter 4 I will present a control approach to drive the activation of the HOG cascade and in Chapter 5 I will show how this pathway can be utilized to control gene expression.

The events following an osmotic shock include a rapid water loss of the cell (see Figure 7). The osmotic up-shift is sensed by the two membrane sensors Sln1 and Sho1 which activate the HOG cascade, thereby leading to the phosphorylation of Hog1 and subsequently to the import of Hog1 into the nucleus within about 3 minutes after an osmotic shock [58–60]. Inside the nucleus Hog1 alters the expression of a vast number of genes [35, 61, 62]. In addition there is a direct stimulation of the enzyme phosphofructo-2-kinase by Hog1 [63, 64], which promotes glycerol production by producing the glycolytic activator fructose-2-6-bisphosphate. Moreover glycerol production is increased by inducing the expression of the glycerol producing enzymes glycerol-3-phosphate dehydrogenase (Gpd1) and glycerol phosphatase (Gpp2). Another mechanism to increase the glycerol concentration is to prevent glycerol from leaking out of the cell. In response to an osmotic shock, the membrane glycerol channel Fps1 closes. This aquaglyceroporin plays a major role in controlling the efflux and uptake of the osmolyte glycerol. Under normal and hypoosmotic conditions Fps1 is open which allows glycerol to leak out of the cell and mutants lacking Fps1 are sensible to hypoosmotic shock [65], while under high osmotic conditions the channel closes and prevents glycerol from leaking out of the cell [32]. After 20-30 minutes the osmotic balance has been restored by glycerol production [66] and Hog1 is exported from the nucleus.

HOG Pathway structure

The HOG pathway is one of the best characterized eukaryotic signaling pathways. While the structure of this pathway is being investigated for 20 years [59], interest has only recently been

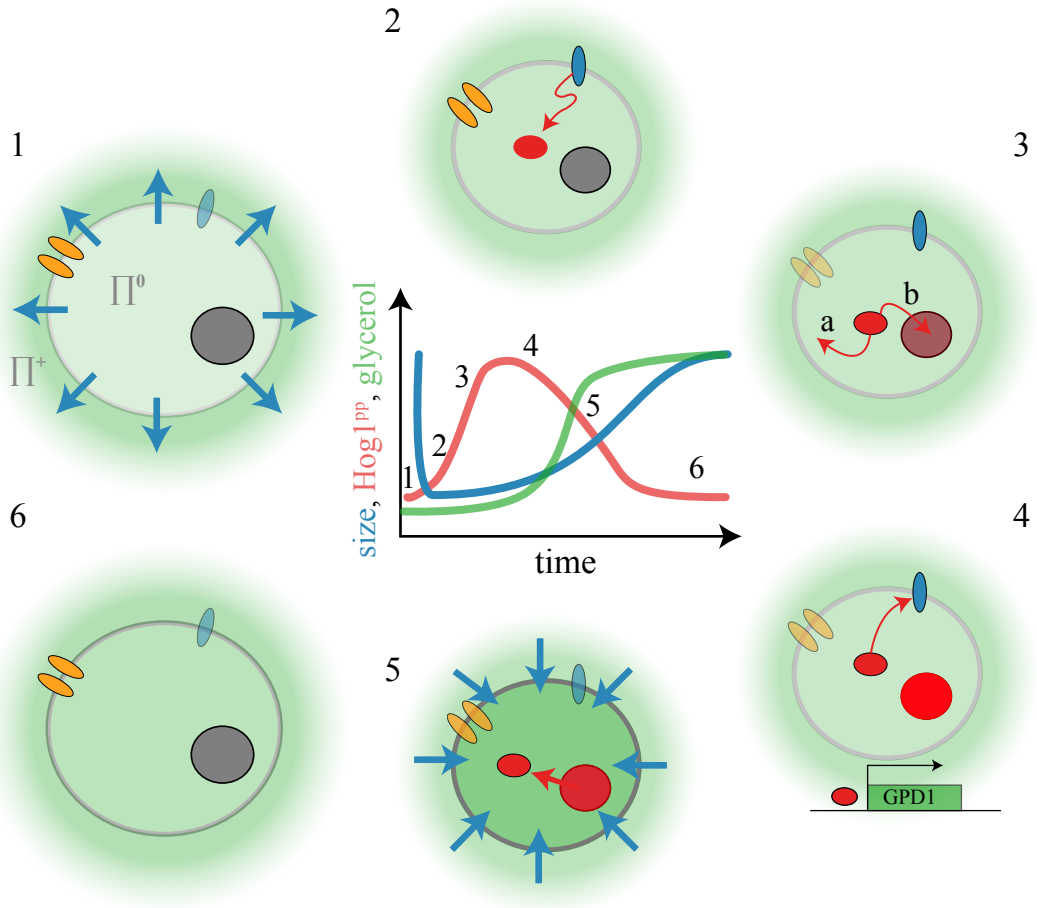


Figure 7: Events following an osmotic shock. The change of the cell size, the Hog1 phosphorylation state and the cellular glycerol level are schematically represented in the center of the image. (1) After an increase of the external osmolarity (green) water flows out of the cell (blue arrows), which leads to a rapid loss in cell size. (2) Membrane receptors (blue) activate the HOG pathway which leads to the phosphorylation of Hog1 (red). (3) Phosphorylated Hog1 performs several actions. (a) The activation of enzymes involved in glycerol synthesis. (b) Translocation to the nucleus. (4) In the nucleus Hog1 induces a large transcriptional response. Glycerol synthesis is increased by inducing the gene *GPD1*, coding for the enzyme glycerol-3-phosphate dehydrogenase. Negative feedback loops lead to the inactivation of the HOG cascade. (5) The increased production of glycerol leads to water influx and the cell size recovers. Hog1 is exported from the nucleus. (6) The cell has restored its osmotic balance. Cell size has recovered so the cell has adapted to the changed environment.

drawn to the dynamical properties of the pathway [66]. The HOG pathway can be grouped into three modules (Figure 8), the Sln1, the Sho1, and a mitogen activated protein (MAP) kinase module. The Sln1 and Sho1 modules are named after their sensory proteins Sln1 and Sho1. Both pathways sense osmotic changes (the exact mechanistic functioning of these activations are unknown) and converge on the signaling protein Pbs2. The Sln1 module is similar to a bacterial two-component system[4] [32]. Because osmotic changes in bacteria are sensed by two-component systems (for example the EnvZ/OmpR system in *E.coli* [67]), it is likely that this branch has evolved from such a system. Sln1 is a transmembrane protein sensing osmotic changes, which is active under ambient conditions and gets inactivated by high osmolarities. The Sln1 branch responds linear up to an external salt concentration of 600 mM [68] and is able to integrate a fast varying signal over time [38]. Cells lacking the Sln1 branch show no response to slightly increased osmolarity and the general response to an osmotic shock is slower as in wild type cells.

The second branch activating Hog1 is the Sho1 branch which is also activated in response to high osmolarity. It is not clear why two parallel branches activate Hog1, in fact the Sho1 branch seems not to be connected to Hog1 in other yeasts [69]. In contrast to the linear response of the Sln1 module, the Sho1 branch is activated in an all or none fashion. The Sho1 branch is not capable of integrating a fast varying signal, and cells lacking this branch still react to osmotic stress in a way similar to wild type cells [38].

The third module of the pathway is made up by a MAP kinase module. MAP kinase modules are signaling cascades involved in many cellular processes such as stress response or the regulation of differentiation and proliferation [66], which are highly conserved from yeasts to humans. They consist of three kinases which phosphorylate each other in a cascade. The final kinase in this chain is termed the mitogen activated protein kinase (MAPK, in this case Hog1) and phosphorylates different effector proteins in response to activation. The MAPK is activated by the upstream MAP kinase kinase (MAPKK, in this case Pbs2), which is in turn activated by a MAP kinase kinase kinase (MAPKKK, in this case Ssk2/22 and Ste11). This means that the Sln1 and Sho1 pathways converge on the MAPKK Pbs2 which then activates Hog1 by phosphorylation. Phosphorylated Hog1 translocates into the nucleus where it activates different transcription fac-

[4]Bacterial two-component systems sense and respond to environmental changes. They consist of a membrane associated sensor and of a response regulator (often a transcription factor).

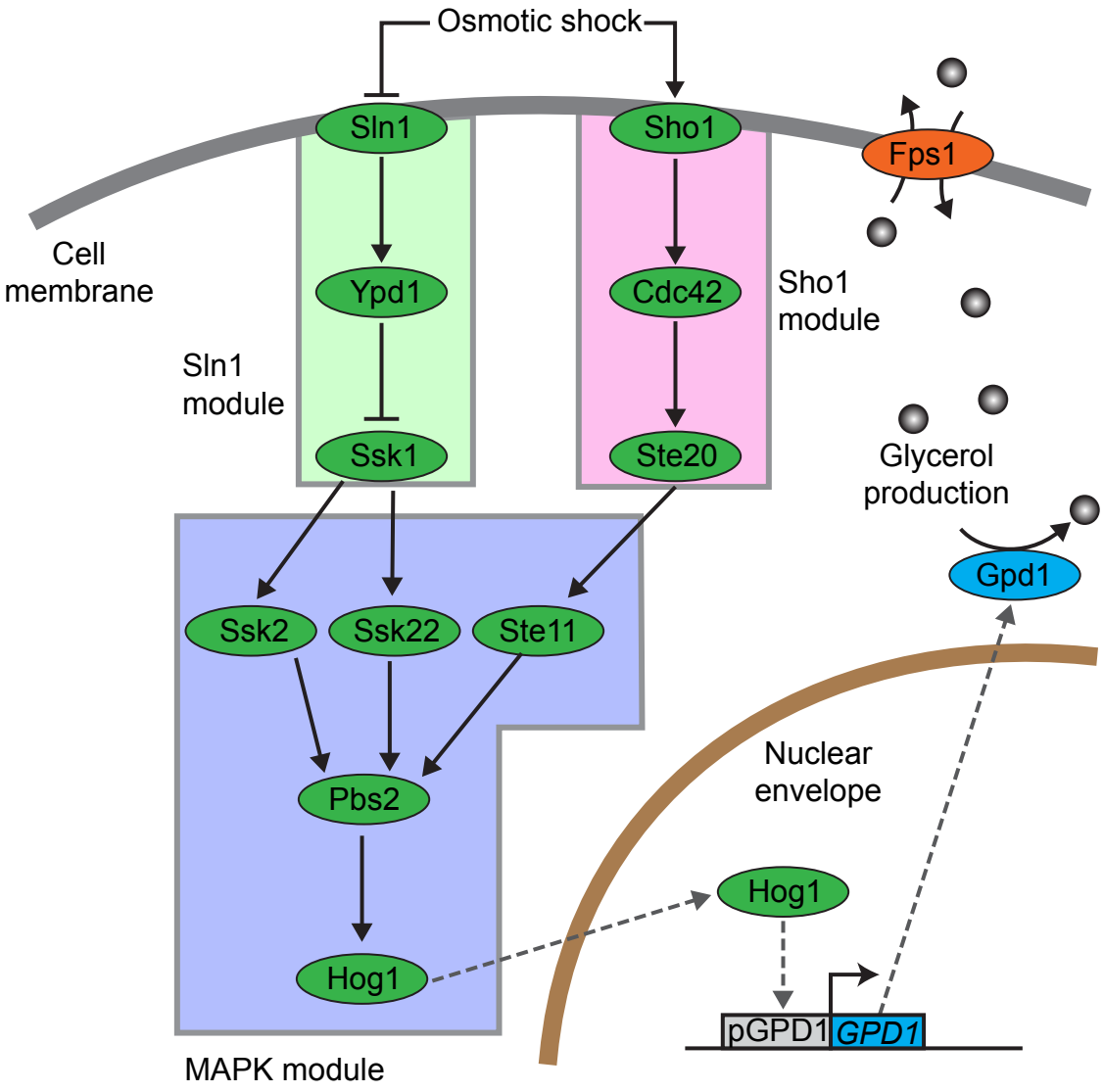


Figure 8: HOG pathway schema. The pathway can be grouped into three modules, the sensory modules Sln1 (green) and Sho1 (pink) both sense osmotic increases and transmit the information to the mitogen activated protein (MAP) kinase module (blue). The two pathways converge on the MAPKK Pbs2 which in turn activates Hog1 by phosphorylation. Phosphorylated Hog1 translocates to the nucleus where it activates the expression of osmoadaptive genes. Adaptation to the hyperosmotic shock is achieved by increasing the intracellular glycerol level. The enzyme glycerol-3-phosphate dehydrogenase (GPD1) synthesizes glycerol, while the membrane glycerol channel Fps1 closes in response to an osmotic shock, which prevents glycerol from leaking out of the cell.

tors by phosphorylation [32].

Feedback

The HOG pathway contains multiple negative feedback loops that operate on different time scales [39] and lead to the inactivation of the pathway after about 20-30 minutes (see Figure 9). One feedback is mediated by the transcriptional activation of glycerol producing enzymes by Hog1. But for a single osmotic shock the time-scale of the pathway adaptation is faster than the time-scale for protein production [39]. In addition it has been shown that adaptation of the pathway requires Hog1 kinase activity, but not protein production ability [70], suggesting that dominating feedback mechanisms are not mediated by transcriptional changes.

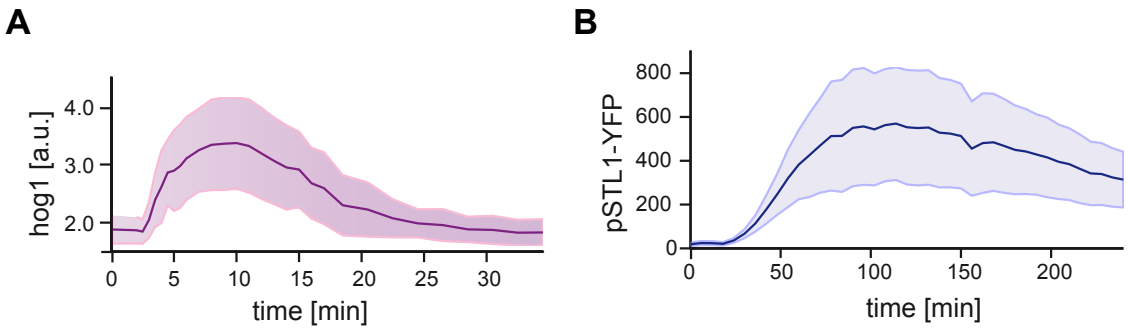


Figure 9: Response of Hog1 nuclear localization and of the osmosensitive gene *STL1* to a step osmotic input. **(A)** Observation of Hog1 nuclear localization after transferring cells from normal media to sorbitol enriched (1M) media in a microfluidic device. The delay of the response is in part due to the delay of the microfluidic device which is between 2 and 3 minutes. After that the pathway reaches maximal activation after about 7 minutes after which Hog1 nuclear localization decreases due to the feedback mechanisms in the pathway. After about 20 minutes pathway activity falls back to the level before the shock, even though the osmotic input is still present. **(B)** Expression of yECitrine driven by the *STL1*promoter following a step osmotic input of 1M sorbitol. The fluorescent protein yECitrine has been placed under the osmoresponsive STL1 promoter. An expression increase can be observed approximately 20 minutes after the onset of the stress. The expression response is transient and reaches its maximum about 100 minutes after the step input with a value of 600 fluorescent units (f.u.).

In addition the pathway shows perfect adaptation, meaning that irrespectively of the strength

of the osmotic shock (as long as it is within physiological bounds), pathway activation will always fall back to its initial activation state after the cell has adapted. This indicates that there exists an integral feedback loop somewhere within the pathway, most likely involving glycerol synthesis activated by Hog1 kinase activity [70]. After the cell has restored its osmotic balance by producing enough glycerol, the pathway is no longer activated, which triggers the export of Hog1 from the nucleus, since different phosphatases are constantly dephosphorylating active Hog1 and other MAPKs [71]. In addition it has also been shown that active Hog1 inactivates Sho1 by phosphorylation [68], which might also contribute to pathway inactivation.

Crosstalk

The osmotic stress response pathway interacts with different cellular pathways like the general stress response pathway or the cell integrity pathway [32]. Different stresses activate in addition to their specific stress response a general stress response. The general stress response pathway enables a cell exposed to a mild type of stress to cope better with a more severe, different type of stress [72]. Stress responses connected by the general stress response include nutrient starvation, oxidative stress, heat shock and hyperosmotic shock. The general stress response is mediated by the transcription factors Msn2 and Msn4, which are in addition to Hog1 activated by high osmolarity. It has been shown that the HOG and the general stress response pathway interact on both signaling and promoter levels [73].

Modeling

The HOG pathway is one of the best studied signaling pathways and serves as an archetypic pathway for eukaryotic MAP kinase pathways, which is illustrated by the vast number of models that have been proposed for this pathway [39, 68, 70, 71, 73–77]. These include detailed kinetic models [71], simple linear models [70] and gene regulatory models [73].

The most detailed model is the one by Klipp *et al.*[71], which consists of 32 ODEs and 70 parameters (see Figure 10A). It takes into account the volume change caused by the osmotic shock. In addition to the Hog1 cascade it also includes gene expression and metabolism modules. Different simplified versions of this model have been proposed, one of them modelling the frequency response of the pathway (see Figure 10B) [74, 75]. Two different linear models have been de-

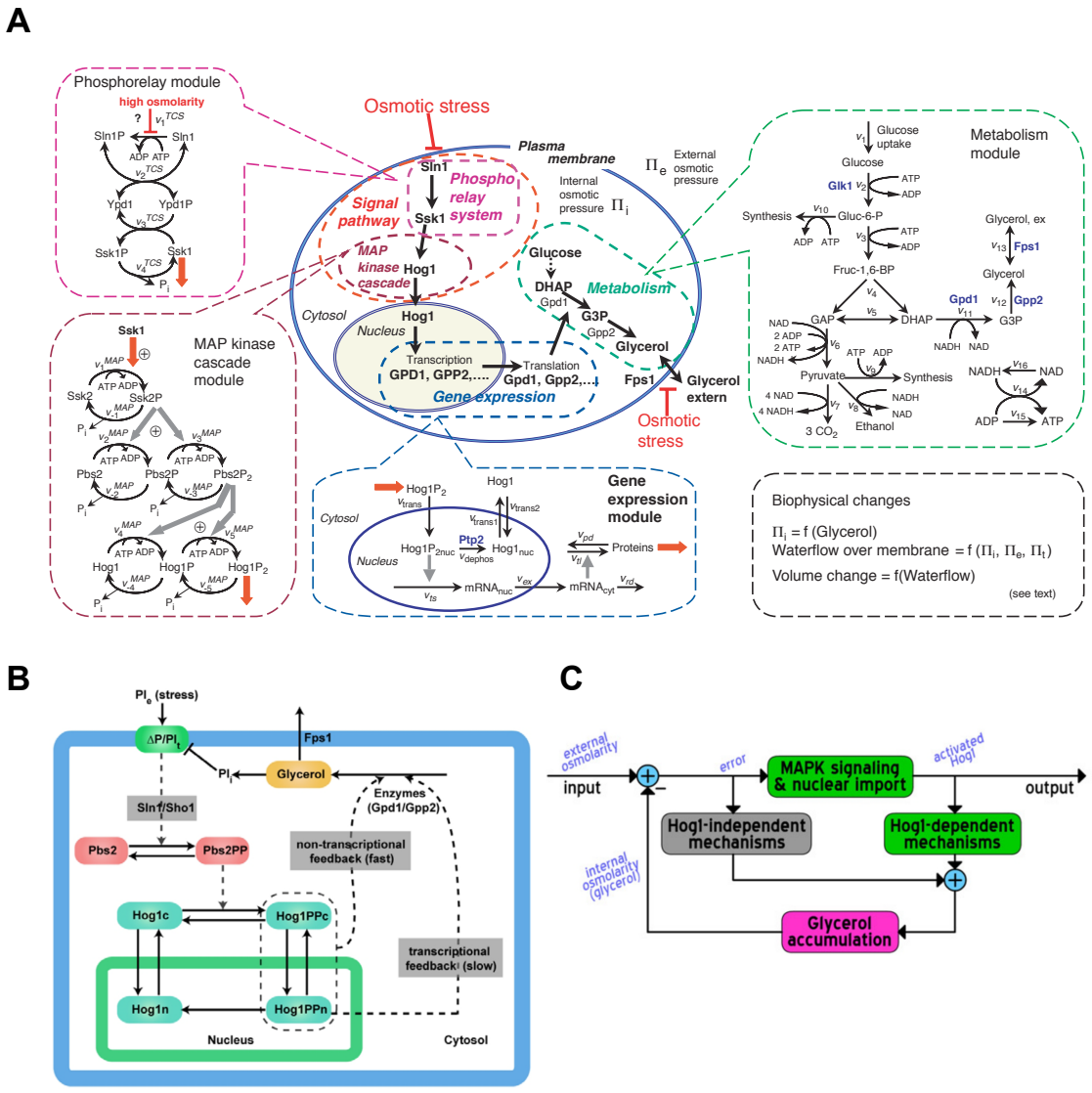


Figure 10: Different models of the HOG pathway. **(A)** Model of Klipp *et al.* [71] including gene expression and metabolic response to HOG pathway activation. The model consists of 32 ODE terms and 70 parameters. **B** Model by Zi *et al.* [75] developed to investigate the response of the HOG cascade to oscillating inputs. **(C)** The minimalistic linear model by Muzzey *et al.* [70] which has been used to identify the dominating feedback mechanism in the HOG cascade. (The figures have been taken from the corresponding publications.)

veloped in the lab of Alexander van Oudenaarden (see *e.g.* Figure 10C) [39, 70]. These models have been developed to analyze the frequency response of the pathway and to study the feedback mechanisms in the HOG pathway.

2.1.4 Control using the HOG cascade

A I have discussed in the previous section, activation of the HOG cascade leads to a major transcriptional response. This means that this pathway can be used, similar to inducible promoters, to induce gene expression. In this thesis I present a gene expression control platform which utilizes the HOG pathway to activate gene expression. One reason for choosing this pathway was the fact that in addition to the gene expression response it is also possible to measure the activation state of the HOG cascade by observing the nuclear localization of Hog1 (see Section 3.1.3). This also allowed me to start the development of the control platform with a controller which drives the activation state of the HOG cascade (see Chapter 4). However the main reason for choosing the HOG pathway and not one of the gene inducible systems described in the next section was to make control challenging. In contrast to synthetic gene inducible systems, the HOG cascade has a fairly complex structure including several feedback mechanisms, which makes it difficult to control this system. Because the gene expression platform I present in this thesis, is not limited to one gene induction system, I will briefly focus on other systems to activate gene expression in the next section.

2.2 Inducible promoters in yeast

Cells alter their gene expression in response to different environmental changes, for example in response to changing nutrient levels. Promoters which control the expression of a gene depending on external signals are called inducible promoters. These promoters are of natural origin, but can often be transferred to other organisms. For the gene expression control I will utilize the HOG cascade as an inducible gene system, but I also give a short review about other methods of gene activation in yeast, because I will later show that the control platform I propose is not limited to use the HOG cascade, but can also control gene expression via other gene induction systems like the methionine inducible promoter (see Chapter 6).

2.2. INDUCIBLE PROMOTERS IN YEAST

Among the most common inducible promoters in *S. cerevisiae* are the *GAL1*, *GAL7* and *GAL10* genes, which are strongly repressed in the presence of glucose and activated in the presence of galactose by the Gal4 activator [78]. Also the promoters of the *CUP1* and the *PHO5* genes which are activated by Cu^{2+}-ions and inorganic phosphate respectively are widely used in yeast [78].

2.2.1 Methionine regulated promoters

The *MET3* and *MET25* genes code for the enzymes ATP sulphurylase and O-acethyl homoserine sulphydrylase respectively, both involved in the synthesis of the amino acid methionine. Both promoters are inhibited by the presence of methionine and can therefore be used as externally repressible promoters. Comparing the two promoters shows that the MET3 promoter is weaker, but also more strictly repressed than the MET25 promoter [79].

2.2.2 Tetracycline promoters

The promoters presented above share the limitation that they are natural promoters in yeast, which limits their use in probing cellular dynamics, because activating these promoters will in addition trigger a natural response in the cell. For example shifting the cells carbon source from glucose to galactose to activate the GAL promoters results in severe changes within the cellular metabolism. In addition, there may exist unwanted upstream effects, for example the *CUP1* gene is not only affected by copper, but also by heat shock, glucose starvation, and oxidation stress [80]. These effects are disadvantageous of course if one wants to study the effect of inducing a protein. At least the problem of an additional natural response can be circumvented by introducing gene activation systems from other species, that are not naturally present in yeast. This independence from the natural cellular pathways is called "otrhogonality" in the synthetic biology community. An example is the tetracycline dependent gene expression system, which stems from Gram-negative bacteria [19, 81–83]. These bacteria developed a resistance to the antibiotics from the tetracycline family, which relies on the efflux protein tetA (A for antiporter[5]), which can pump tetracycline out of the cell [84]. tetA has negative effects on the cell growth and viability of *E. coli* [85] and is only expressed in the presence of tetracycline or one of its relatives. This mechanism works by the expression of a tet repressor protein (tetR), which binds the promoter

[5]Similar to a symporter, an antiporter is a membrane transporter which transports two different molecules, but opposed to the symporter not in the same, but in opposite directions.

of tetA and its own promoter, thereby repressing expression of itself and of tetA. Binding of tetracycline to tetR lowers its affinity to the tetO operator, wherefore the genes can be expressed in the presence of tetracycline.

Tet-off system

Early studies using the tetR repressor protein in eukaryotic cells placed the tetO operator sequence within the promoter of a target gene, which enabled the repression of the promoter in the absence of tetracycline [86–88]. In 1992 Gossen and Bujard constructed the Tet-off system, by fusing the tetR protein with the activating domain of viron protein 16 (VP16) from the herpes simplex virus [89] (see Figure 11A). This generated the tetracycline controlled transactivator tTA, which activates transcription in the absence of tetracycline on a minimal promoter derived from the human cytomegalovirus promoter IE combined with tet operator sequences. To use this system it is necessary to place the gene to be influenced under the control of the tet promoter and to constitutively express tTA. These constructs have been developed for human HeLa cell lines, but they function also in *S. cerevisiae* [90].

Tet-on system

In practice the Tet-off system has several drawbacks. One is the need of sustained tetracycline presence to maintain an uninduced expression state, because excessive exposure of cells to tetracycline might have side effects. Also, once tetracycline has been added to the media, it is not always easy to remove it, because exchanging the cellular media requires additional treatments like centrifugation or advanced setups like microfluidic devices. To develop a system which can be switched on in the presence of tetracycline, a random mutagenesis of the tetR protein has been done, to find a variant of tetR that can bind the tetO operator only in the presence of tetracycline [91]. This resulted in finding the reverse tet repressor (rTetR), which shows the desired binding characteristics. Fusing rTetR with VP16 resulted in the reverse transcriptional transactivator (rtTA). rtTA induces gene expression in the presence of tetracycline. In addition rtTA shows a better response to the tetracycline analog doxycyline than to tetracycline [91].

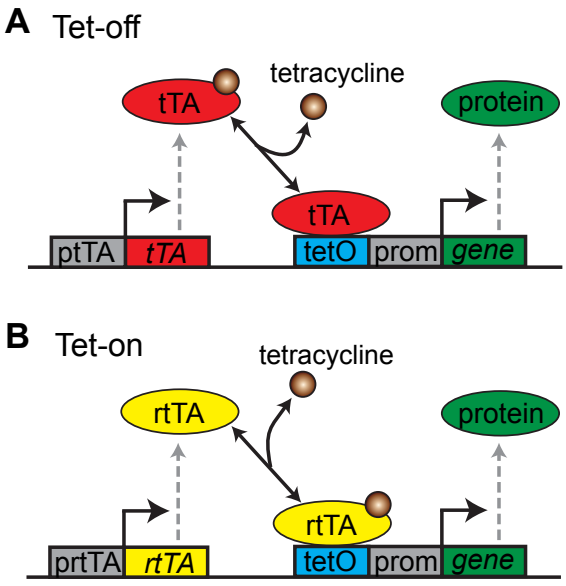


Figure 11: Tetracycline gene induction systems. **(A)** Tet-off system. The tetracycline controlled transactivator (tTA) binds to the tet operator sequence (tetO) if in no tetracycline is bound. This activates the expression the controlled gene. If tetracycline is bound, tTA dissociates from tetO, so the gene is not expressed. **(B)** Tet-on system. The reverse tetracycline controlled transactivator (rtTA) binds to tetO in the presence of tetracycline, thereby activating the expression of the target gene.

Improvements

The Tet-on system has been further improved by random mutagenesis. This resulted in a system with lower basal expression and an increased doxycycline sensitivity [92]. In addition the sequence of the rtTA has been optimized for expression in human cell lines. An improved system for yeast has been proposed by Bellí *et al.*, which uses transcriptional repressors fused to TetR or respectively rTetR to decrease basal expression [93]. This repressor prevents gene expression when tTA or respectively rtTA are not bound, thereby reducing the basal expression level. Nevozhay *et al.* engineered a tetracycline inducible gene activation system with a linear dose response curve, by introducing a negative feedback loop acting on the expression of the tetR repressor [94].

2.2.3 Optogenetic control of cellular activities

All chemical compounds that are being used to alter cellular behavior share the limitation that the timescales on which they act are rather slow because they need to be added to the media, diffuse into the cell and perform their function within the cell. This limitation does not apply to the recently emerging field of optogenetics, in which light-sensitive proteins are used to control cellular processes. The name optogenetics stems from the fact that these systems react to light (opto) and are encoded genetically. Because the temporal and spatial resolution with which light can be applied to cells or cellular regions excels that of any previous method, optogenetic methods are the first choice to control fast cellular processes like signaling, or events within sub-cellular regions or single cells. Optogenetic techniques are based on a light-induced conformation change of a protein which can stem from light-sensitive bacteria [95], from plants [18], or animals [96]. By using these light-sensitive proteins directly [97] or by combining them with signaling receptors [98] it has become possible to act on cellular systems by using light with a proper wavelength.

So far, the most prominent use of optogenetics has been in the field of neuroscience, because stimulating neurons requires a high temporal precision which is, at least in combination with the cellular specificity optogenetic techniques offer, currently not achievable with other techniques [99]. The first genetically encoded control of neuronal activity was published by Zemelman *et al.* in 2002 [96], who fused a *Drosophila* photoreceptor to the α-subunit of a large G-protein in order to control the signaling activity in rat neurons *in vitro*. Other approaches used optically gated ion channels originating from green algae [97, 100] or from Archaea [101] to stimulate or inhibit neurons. These developments gave researchers the ability to control the neuronal activity even in live *Drosophila*, which enabled them to trigger specific behaviors like jumping or wing beating in fruit flies [102]. By stimulating arousal associated neurons, Adamantidis *et al.* managed to wake up sleeping mice using an optogenetic control [103]. Han *et al.* showed that an optogenetic control of neurons is feasible in primates [104].

Apart from controlling neuronal activity, optogenetic techniques offer also unprecedented opportunities to act on cellular signaling dynamics. A tool that has been applied in many approaches controlling cellular signaling is the light based association of phytochromes (phy) and phytochrome interacting factors (PIF) from *Arabidopsis thalina* [105]. Phy can switch between two conformations in response to red or infrared light and the conformation resulting from exposure to red light binds PIF. Exposure to infrared light releases the binding between phy and

PIF. This tool has been used to control gene expression in yeast cells by fusing phy to the GAL4-binding-domain (phy-GBD) and PIF the GAL4 activation domain (PIF-GAD) [18].

Especially for the control of cellular signaling, where precise timing is important optogenetic techniques are likely to be more widely used in the future, but in order to make them widely accessible by the scientific community, light delivery protocols as well as possibilities to read out the activation state have to be improved [106].

2.3 Control theory

Control theory is a field in mathematics and engineering that studies dynamical systems whose behavior can be influenced externally. Examples of control theory applications range from simple temperature control systems (*e.g.* air conditioning) to complex safety critical digital controllers like the autopilot of an aircraft. If the controller monitors one or several outputs of the system we call it a closed-loop controller (feedback control), if not its called an open loop controller (non-feedback control). A schematic representation of open- and closed-loop controllers is shown in Figure 3 in Section 1.1.5. In the following I will briefly present the basic concepts of this vast field.

2.3.1 Dynamical systems

Control theory is concerned with the behavior of dynamical systems. A dynamical system is a system whose internal state is changing in time. Consider for example the growth of a bacterial population. We can define the number of cells at a certain time point t as the variable x_t. All values that x_t can attain are integral and define the state space, in this case $x_t \in \mathbb{N}$. If the state space is discrete we speak of a discrete dynamical system, if its continuous, of a continuous dynamical system. The dynamics of the system are said to be deterministic if there exists a unique solution for the temporal evolution of the system. Otherwise the system is non-deterministic. Imagine now that each cell will duplicate at each time step and that cells never die. In this case the temporal evolution of the dynamical system can be written as

$$x_t = 2x_{t-1} \tag{2.1}$$

with the initial condition $x_0 \in \mathbb{N}$. The solution of this initial value problem is

$$x_t = 2^t x_0 \tag{2.2}$$

In this case we are dealing with a discrete-time system, but we can also define a dynamical system for continuous-time. For example if we approximate the number of cells with a continuous variable $y \in \mathbb{R}^+$ we can model the bacterial growth with the following differential equation

$$\dot{y}(t) = \frac{dy(t)}{dt} = \mu y(t) \tag{2.3}$$

with growth rate $\mu \in \mathbb{R}^+$ and initial condition $y(0) = y_0$. The solution in the continuous case is

$$y(t) = e^{\mu t} y_0 \tag{2.4}$$

Dynamical systems theory plays an important role in biology to model different dynamical processes like growing cell populations, signaling or metabolic pathways, genetic networks or the spreading of a drug in a body. These models can help to understand the dynamics of the underlying processes and often lead to novel biological hypothesis [107, 108]. In addition to such predictive models, which provide generally an abstracted representation of the modeled process, large scale descriptive models integrate biological knowledge [109, 110]. While these models usually did not include any dynamics but instead focused on the structure of biological systems, there are efforts to combine these two modeling approaches for example by developing whole-cell dynamical models [111].

The mathematical formalism of the models used in biology varies depending on the application and on the available data. To review the different formalisms would be beyond the scope of this work. The models I will use in this work are ordinary differential equation (ODE) models, ordinary meaning that all derivatives are with respect to one variable, in this case time. The assumptions that are being made when using ODE models to describe biomolecular processes are (i) that the described system is well stirred and (ii) that the molecule numbers within the volume which is considered (*e.g.* concentrations) are high enough to allow smooth local averaging.

2.3.2 Feedback control

Dynamical systems theory is of vital importance in control because it allow to predict how a system behaves. Model-based control approaches use a model of the controlled system, which is used to find optimal control inputs. Note that the optimality of the control inputs depend on the specific use case, because different control applications have different requirements, for example minimizing the energy consumption or the speed of convergence. Even for controllers that do not rely on a model (model-independent control), a model is usually used in the development of the control system to tune the controller.

In the following I will describe the two controllers which I will use in this work, the proportional-integral-derivative (PID) controller and the model predictive controller (MPC).

PID control

The most common feedback controller is the proportional-integral-derivative (PID) controller [112]. It works by calculating the difference between the observed and desired state of the system, which is called error ($e(t)$). The control $u(t)$ which is applied at a certain time t is then just the weighted sum of a term proportional to the current error, a term integrating the current error over time and a time derivative of the current error.

$$u(t) = k_P \cdot e(t) + k_I \cdot \int_0^t e(\tau)d\tau + k_D \cdot \frac{d}{dt}e(t) \tag{2.5}$$

The factors k_p, k_I and k_D are the weights for the three different terms. The proportional term k_P quantifies the direct influence that the error has on the control, while the integral term k_I can be thought of as the "memory" of the controller, because it quantifies the effect of past errors on the control. High values for the derivative term k_D result in a fast responding controller and the derivative term can be thought of as a prediction of how the system will react in the near future. For noisy systems, the derivative term is often neglected ($k_D = 0$), which yields a proportional-integral (PI) controller. A PID controller does not require any structural model about the controlled system and is widely used in practice due to its simple form. In order to construct an effective controller that converges fast to the target value, without showing large oscillations around the target, the weight factors have to be chosen carefully. If a model of the controlled system is available, this can be used to help tuning the parameters based on *in silico* analysis.

Model predictive control

A model predictive controller (MPC) makes use of a model of the controlled system to simulate the response of the system to various control inputs. MPC relies on a cost function which defines the optimal control strategy and that is based on the difference between a desired output profile of the controlled system and the prediction of the model for a certain input. This cost function can in addition include other measures, for example a term ensuring that the control is achieved with a minimal consumption of energy. After each observation the current state of the system is estimated (see Section 2.3.3). Starting from this estimated state, an optimal control strategy is identified by minimizing the cost function over a (short) time horizon. This strategy is then applied for a short while, until a new observation is made and the process is repeated (receding horizon strategy). MPC works with almost any kind of model provided that a cost function can be defined for each possible input and efficiently minimized. Moreover, it allows the integration of complex constraints on the control values.

2.3.3 State estimation

To simulate the behavior of the controlled system, the MPC requires knowledge of the current state of the model used to describe the system. Two problems arise when estimating this state based on observations. Firstly, observations can be noisy which leads to a discrepancy between observed and real values. Secondly, observations can be partial meaning that information is only available on a limited set of the state variables. To estimate the states of the unobserved (hidden) variables, we can exploit the influences that the hidden variables have on the observed ones. The problem of reconstructing the full state of a model based on a series of noisy observations of a part of the state space is called state estimation problem. This state reconstruction is not possible for all systems (for details see [112]).

Kalman filter

A Kalman filter is a process to estimate the state of a linear dynamical system based on a series of (noisy) measurements of a subset of the variables of the system [113]. It works in a two step process. In a first prediction step the current state of the system and its uncertainty are estimated using the model of the system and previous state estimates. Next, in the update step, the esti-

mated state is updated based on observations and a weighted average between the computed and observed state is chosen.

The Kalman filter assumes a time-discrete model of the form

$$x_k = Ax_{k-1} + Bu_k + w_k \tag{2.6}$$

where u_k denotes the control which is applied at time-step k, $w_k \sim N(0, Q)$ is the process noise which is assumed to be Gaussian with zero mean and covariance Q, and A and B are matrices. We are not able to observe the full state of x and the observations we can make are noisy wherefore each measurement z_k has the following form

$$z_k = Hx_k + v_k \tag{2.7}$$

where the matrix H describes the observation model, which maps the state of the system to the observed state and $v_k \sim N(0, R)$ is the measurement noise of the system which is again assumed to be Gaussian with zero mean. We define now $\hat{x}_{k|l}$ and $P_{k|l}$ as the estimated state of the system and the covariance of this state estimate at time k using observations up to time l. In the prediction step, the Kalman filter estimates the state and its covariance based on the model description.

$$\hat{x}_{k|k-1} = A\hat{x}_{k-1|k-1} + Bu_{k-1} \tag{2.8}$$

$$P_{k|k-1} = AP_{k-1|k-1}A^T + Q \tag{2.9}$$

In this form the filter just simulates the model, without taking observations into account. Observations are integrated in the update step:

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k(z_k - H\hat{x}_{k|k-1}) \tag{2.10}$$

$$P_{k|k} = P_{k|k-1} - K_kHP_{k|k-1} \tag{2.11}$$

with Kalman gain K_k, which describes the weight that is given to the observation and is defined as

$$K_k = P_{k|k-1}H^T(HP_{k|k-1}H^T + R)^{-1} \tag{2.12}$$

The Kalman filter has been developed for linear systems, but an extension to nonlinear systems is possible by linearizing the nonlinear model around its operating points [113]. In practice, in the nonlinear case equation 2.6 is replaced by

$$x_k = f(x_{k-1}, u_{k-1}) + w_{k-1} \tag{2.13}$$

and the transition matrix A is approximated by the Jacobian of f

$$F_{k-1} = \left.\frac{\partial f}{\partial x}\right|_{\hat{x}_{k-1|k-1}, u_{k-1}} \tag{2.14}$$

2.4 Global optimization

An optimization problem is the problem of finding an input which minimizes (or equivalently maximizes) a certain function. In addition the value of the input variables often cannot be chosen freely, such that only certain inputs are allowed (constrained optimization problem). The algorithms used to solve optimization problems can be grouped into local and global optimization methods. A local optimum of a cost function is a point for which the value of the cost function is lower (or higher for a maximization problem) than for all points in a given neighborhood. In many non-linear optimization problems, the cost function has many local optima but one is usually interested in finding the global optimum, that is the best of all local optima. Local optimization methods often rely on local properties of the cost function, for example on its gradient. These methods converge fast but are often "trapped" in local optima. In contrast to local optimization methods, global search methods aim at finding the global optimum of an objective function. The covariance matrix adaptation evolutionary strategy (CMAES) [114] is an evolutionary global optimization algorithm. Evolutionary search methods belong to the class of stochastic optimization methods, which are in general well suited to solve global optimization problems, because in contrast to deterministic search methods they allow the search path to step out of local minima. Evolutionary optimization algorithms apply ideas originating from the natural selection theory in biology to optimization problems. In general most evolutionary strategies work with a random population of solutions, from which the best ones are selected and recombined to a new generation of solutions. The evaluation of the solutions is based on the rank of their corresponding

fitness value rather than on the absolute value.

In CMAES candidate solutions are sampled according to a multivariate normal distribution

$$x_i \sim m + \sigma N_i(0, C) \tag{2.15}$$

where m denotes the mean of the population, σ the step length and C the covariance matrix describing the shape of the search problem. Mean and covariance are selected here following a maximum likelihood principle. The mean m is selected such that the likelihood of previously successful candidate solutions is maximized. Also the covariance matrix is selected in such a manner that the likelihood of previous selected steps that were successful is increased.

CMAES is often applied in systems biology to search for parameters of ODE models [115–117]. In the context of this work I will use CMAES to search for optimal control strategies.

3

Materials and Methods

In the previous chapter, I have explained the background of the work presented here, by introducing the biological system I control and by giving a broad overview about control methods. In this chapter I will focus on the specific methods I have used to develop the control platform. These include methods for cell tracking and automated quantification of fluorescent measurements that I have implemented in MATLAB. In addition I describe the microfluidic devices I have used in this work.

3.1 Image analysis and cell tracking

A feedback controller using a microscope as a sensor requires the automatic analysis of microscopy images and this analysis needs to be fast and robust to meet the real-time requirements. For the control of single cells we also need to be able to identify single cells in an image and to track them over time.

3.1.1 Detection of yeast cells in images

The quantification of fluorescent markers requires knowledge of the position and size of each cell in the image. It is possible to automatically detect cells within an image with different methods [118]. The difficulty of this detection is often to separate adjacent cells from each other without over-segmenting the image. Yeast cells have a well defined quasi round shape, a feature that can be exploited to detect them in an image.

Hough transform

The Hough transform [119] is a method to detect parametrizable objects such as lines or circles in images. The parameter space is here discretized by an accumulator array. Each entry of the accumulator array represents a small region in parameter space and is initialized with a zero value. For example a line in an image can be described by its angle and its distance from a point of origin. In this case the accumulator array would be 2-dimensional and each entry in the array would correspond to a certain line in the image.

The algorithm then first detects edges in the image, for example by thresholding a gradient of the image. For each pixel in the image that belongs to an edge, the corresponding values in the accumulator array are then incremented. For example for line detection all values in the accumulator array that represent a line going through that point are incremented. It is important that all pixels which belong to the same object (*e.g.* which are colinear in the example of line detection) share a cell in the accumulator array. This has to be ensured by a careful choice of the parametrization of the object to be detected. Instances of the detected objects are then identified by finding local maxima in the accumulator array. An example of a Hough transform is depicted in Figure 12. This example explains the Hough transform for line detection, but the Hough transform can easily be extended to detect circles [120]. In this case we are dealing with a 3-dimensional accumulator array (radius, x-position and y-position).

Other cell detection methods

The Hough transform is not the only algorithm suitable for segmenting images of cells, but I chose it in this context because it works very well on images of yeast cells by exploiting the feature of their round shape [121]. Other methods for segmenting images are for example thresholding the

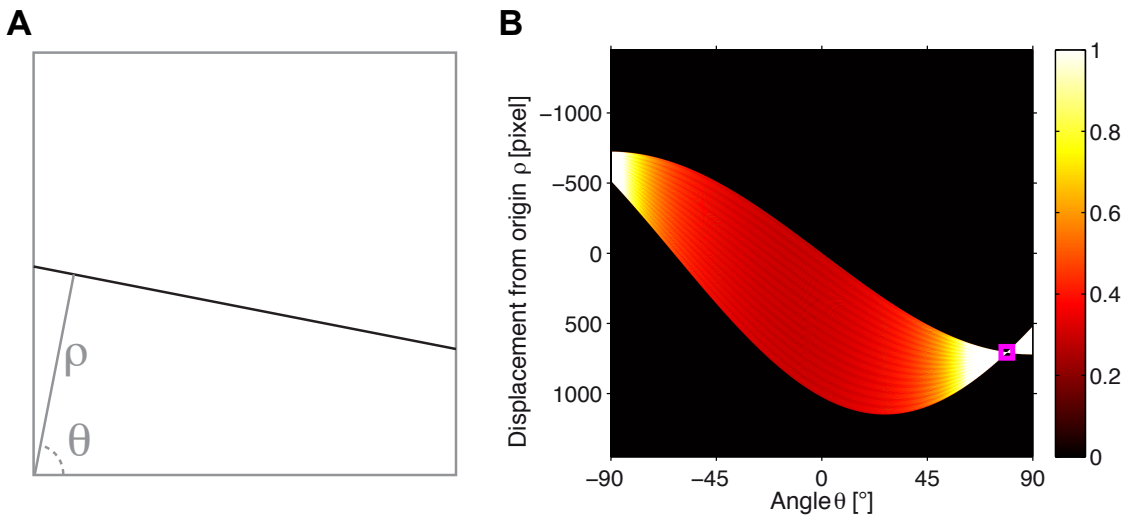


Figure 12: Example of a Hough transform to detect lines. **(A)** A line within an image can be parametrized by the displacement ρ between the line and the point of origin and the angle Θ between a vector orthogonal to the line and the x-axis. **(B)** Accumulator array in which each entry corresponds to a line in the image. The values in this array indicate how often an edge was identified in the image which lies on the line corresponding to the position in the accumulator array. The pink square denotes the maximum of the accumulator array, so the position of this maximum in the array gives the parameters of the detected line in **(A)**.

image to detect objects that are brighter than the background and the subsequent identification of connected objects in the threshold image. This approach does not work well if the cells occur in clusters, because the thresholding method cannot distinguish the single cells in a cluster. Another very common method is the watershed algorithm, which regards an intensity image as a landscape and segments the image based on the height of the landscape by continuously flowing water outgoing from local minima. A problem of the watershed method is that it often leads to oversegmentation of an image, which requires additional treatment of either the segmentation result or of the original image. A comparison of the watershed method to the Hough transform is shown in Figure 13. Often the thresholding method is combined with the watershed method, by converting the binary threshold image to a landscape by applying a distance transform [122]. A review about different methods to segment images of cells can be found in [118].

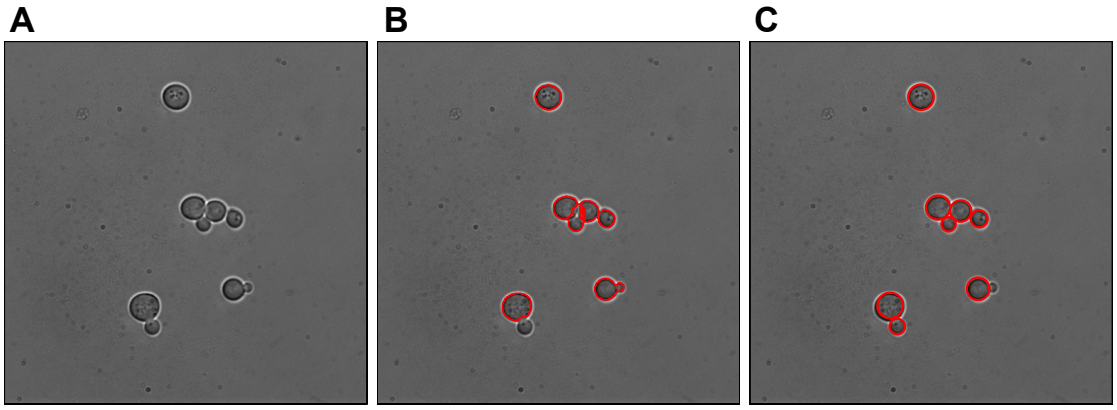


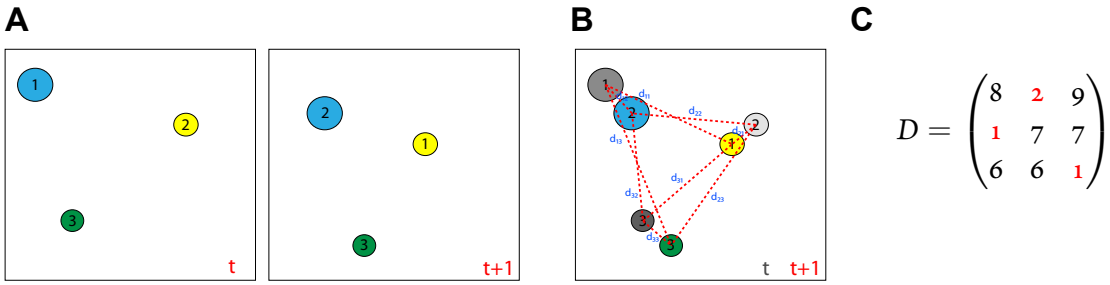
Figure 13: Comparison of watershed and Hough transform to detect yeast cells. **(A)** Bright field image of *S. cerevisiae* cells. **(B)** The watershed transform to detect cell boundaries often leads to over-segmentation. **(C)** A circular Hough transform approximates the cellular boundaries with circles.

3.1.2 Celltracking

Once the positions of the cells at different time points are known, the cells in the different images have to be matched. This means that each cell in an image at a certain time-point has to be identified with a cell from the previous image. This is not a trivial problem since the number of cells in two consecutive images is not necessarily the same. Cells can disappear because they move out of the field of view, or cells can appear by moving in the field of view or by cell-division. The problem can be solved by computing the distances between all cells in one image and all cells in the consecutive image. Then each cell from the later image has to be identified with a cell from the previous image, minimizing the sum of the corresponding distances (see Figure 14).

Problem formulation

The problem of identifying the same cells in two consecutive images with the help of a cell-to-cell distance matrix can be formulated as a binary integer linear programming (BIP) problem. A BIP



$$D = \begin{pmatrix} 8 & 2 & 9 \\ 1 & 7 & 7 \\ 6 & 6 & 1 \end{pmatrix}$$

Figure 14: Celltracking problem. **(A)** Illustration of two consecutive images of three cells. The order in which the cells are detected may change, which is indicated by a change in numbering. **(B)** Pairwise distances of the cells in the consecutive images. **(C)** Distance matrix corresponding to the distances in (B). Minimal distances are denoted in red.

problem can be expressed in the following form:

$$\min_x c^T x \tag{3.1}$$

such that

$$Ax \leq b$$
$$A^{eq} x = b^{eq}$$
$$x \in \{0,1\}^n$$

with the cost vector $c \in \mathbb{R}^{+n}$, the inequality constraint matrix $A \in \mathbb{R}^{m \times n}$ and the equality constraint matrix $A^{eq} \in \mathbb{R}^{k \times n}$. If we fill the cost vector c with the entries of the cell-to-cell distance matrix we can interpret the solution vector x as a mapping in which pairs of cells have been identified as identical. The vector x has one entry for each pair of cells, a 1 in x meaning that the corresponding cell-to-cell distance is minimal, so the two cells are assumed to be identical. The inequality constraint matrix A and the vector b can be chosen in a way that each cell from the current image is only assigned to one cell from the previous image and the other way round (see Figure 15). The equality constraint ensures that the maximal number of cells is mapped. The advantage of formulating this problem as a BIP problem is that for such problems efficient solving

algorithms like the simplex algorithm [123] have been developed[1].

A

$$D = \begin{pmatrix} 8 & 2 & 9 \\ 1 & 7 & 7 \\ 6 & 6 & 1 \end{pmatrix} \qquad \Rightarrow \qquad c^T = \begin{pmatrix} 8 & 2 & 9 & 1 & 7 & 7 & 6 & 6 & 1 \end{pmatrix}$$

B

$$A = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \qquad b = \vec{1}$$

C

$$A^{eq} = \vec{1} \qquad b^{eq} = \min(\#\text{old cells}, \#\text{new cells})$$

D

$$x^T = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Figure 15: Example of linear constraint mapping. **(A)** The weight vector c is constructed by appending the rows of the distance matrix D. **(B)** The inequality constraint matrix A and the vector b ensure that only one entry per column and row of the distance matrix are selected. **(C)** The number of ones in x has to be equal to the number of cells that can be mapped. **(D)** Solution vector x corresponding to the optimal solution of this example.

3.1.3 Quantification of Hog1 nuclear enrichment

The activation of the HOG pathway leads to the import of the Hog1 protein into the nucleus where it influences the expression of various genes (see Section 2.1.3). It has been shown that the nuclear localization of Hog1 correlates with its activation (i.e. phosphorylation), wherefore we can measure the activation state of the HOG pathway by measuring the nuclear localization of Hog1 [38]. In this work I will use strains in which the Hog1 protein has been fused to a fluorescent protein, so its localization can be observed (see Figure 16). Depending on whether in

[1]The tracking software I developed can either use the native matlab function `bintprog()` to solve the BIP problem, or alternatively the open source software lp_solve (http://lpsolve.sourceforge.net) which shows a significant performance increase over the native matlab implementation.

addition a fluorescent label for the nucleus is available, different image analysis techniques are worth considering.

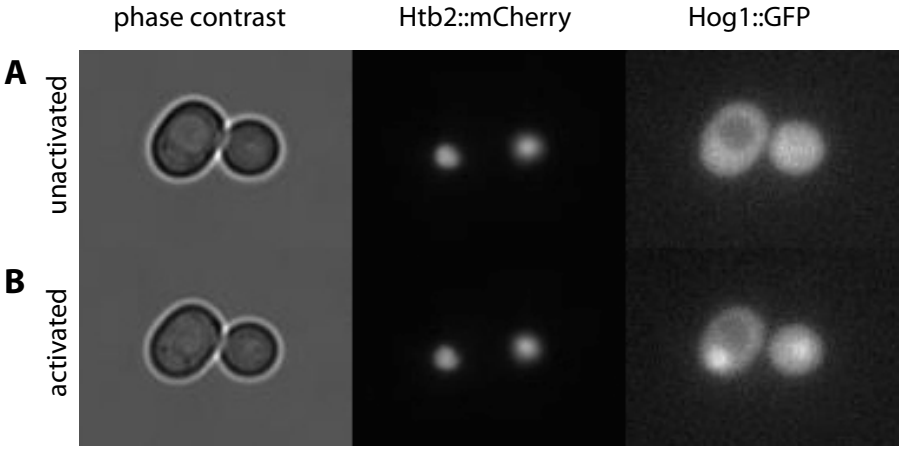


Figure 16: Example of Hog1 nuclear localization in response to an osmotic shock. Images show normal phase contrast, the nuclear marker Htb2 tagged with mCherry and Hog1 labeled with GFP. Under normal conditions **(A)** Hog1 is cytoplasmic, while it translocates to the nucleus after an osmotic shock **(B)**.

Colocalization with nuclear marker

If both Hog1 and a nuclear marker carry a distinct fluorescent label, we can define the activation state of the HOG pathway by the colocalization of Hog1 and the nuclear marker [38]. The relative Hog1 nuclear colocalization can be defined as the ratio of the mean pixel intensity of the Hog1 marker inside the nucleus and in the cytoplasm.

$$h(t) = \frac{\langle \text{Pixel intensity} \rangle_{nuc}}{\langle \text{Pixel intensity} \rangle_{cyt}} \tag{3.2}$$

The area of the nucleus is here detected by thresholding the fluorescent image of the nuclear marker. The cellular area can be detected either with the Hough transform described in Section 3.1.1, or by thresholding a cytoplasmic fluorescent marker. A way to normalize the colocalization is to divide it by the colocalization value measured in the non-activated state.

Contrast of the cell-image

In case a nuclear marker is not available, the activation state of the HOG cascade can be quantified by the contrast of the Hog1 marker inside the cell. In practice I defined this contrast as the difference between the maximal and the minimal fluorescent intensity inside the cell. To smooth this measurement, the maximal and minimal intensities are computed by averaging the 15 brightest and 15 dimmest pixels respectively. In order to prevent pixels that do not belong to the cellular area are taken into account, only pixels whose distance to the center of the cell is less than 60% of the cell radius are taken into account. A comparison of the colocalization and contrast methods in quantifying Hog1 localization data (see Figure 17) shows that the two methods yield largely similar results.

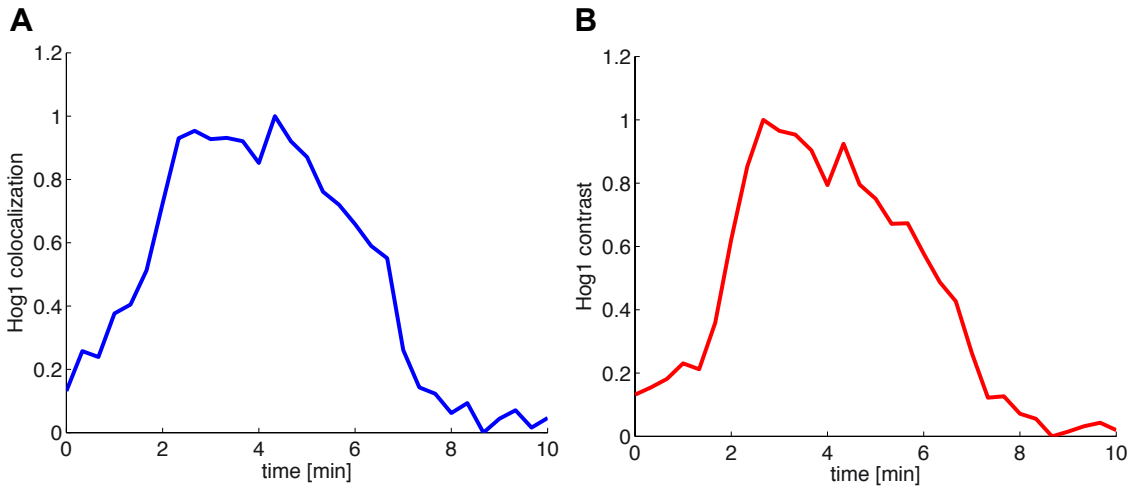


Figure 17: Comparison of colocalization and contrast to quantify Hog1 nuclear localization. Nuclear localization of Hog1 in response to a 6 minute long osmotic shock starting at time 1 minute, quantified either by colocalization **(A)** or by contrast **(B)**.

3.1.4 Quantification of gene expression using a cytoplasmic fluorescent marker

In this work I use a cell-line in which the *STL1* gene has been replaced by the fluorescent marker yECitrine. The quantification of this cytoplasmic osmo-stress responsive marker is done by quantifying the mean pixel intensity inside each cell and subsequent subtraction of the fluorescent

background level. After detection of the cellular boundaries using the Hough transform this quantification is straightforward. One problem is that the cell-size of the cells changes in response to an osmotic shock. Since the amount of fluorescent proteins does not change at the same time, this means that the detected fluorescent intensity increases with the shrinking cell-size. In this work I neglect this phenomenon. The increase of fluorescent intensity in response to a cell-size change caused by 1M sorbitol is on average around 12 percent. Given that the average coefficient of variation for each cell in a control experiment is around 20 (see Figure 35), neglecting this phenomenon will most likely not have a large impact on the experimental results.

3.2 Microfluidics

In addition to observing and tracking cells, a closed loop control platform requires to be able to act on the cellular environment. This is possible with microfluidic devices, which are small devices made out of glass or synthetic polymers that allow the manipulation of small (nanoliter scale) volumes. Their low production cost and small reagent consumption due to their small size led to many successful applications in biology in the recent years. On the one hand there are biotechnological applications like fluorescence activated cell sorting (FACS) [124], nanovolume polymerase chain reaction (PCR) [125] or oligonucleotide synthesis [126]. On the other hand microfluidic devices led to a great advancement in the observation of cells while being able to precisely manipulate their environment, especially in combination with fluorescent imaging techniques. They allow to observe single cells for many generations while the external conditions can be altered within minutes or even seconds. This permitted to better understand signaling dynamics of cells by stimulating pathways in a precise, well defined manner [38, 39, 127]. The possibility to generate for example oscillatory inputs provides a completely novel way to discover the functioning of genetic networks [37, 128]. In addition microfluidic devices can be used as a microchemostat, meaning that environmental conditions can be kept constant over a long period, because possible metabolic products released by the cells are washed away.

To construct microfluidic devices, the same techniques used for the construction of integrated circuits on a silicon wafer are used. In a first step a master is created using photolithography. The process of photolithography is the transfer of a motif from an optical mask to a photoresist which covers the surface of a wafer. A photoresist is a chemical for which the solubility in a certain solution (developer) depends on whether the photoresist has been exposed to light or not.

Simply speaking, a positive photoresist becomes soluble in response to light, a negative photoresist becomes insoluble in response to light. To transfer the motif from the optical mask to the photoresist, the two are placed in contact and exposed to intense light. Either the non-exposed part (negative photoresist), or the exposed part (positive photoresist) can then be washed away by the developer.

To control the thickness of the photoresist, which later defines the height of the microfluidic chamber, a certain amount of photoresist is applied to the wafer, which is then spun at a certain speed for a defined time in a spin-coater. Another control of the thickness is the viscosity of the photoresist. Before and after the exposure, the wafer has to be heated to a certain temperature for a specific time. In a last step the surface of the pattern is treated with trichloro(tridecafluoro-octyl)silane in a process called silanization to facilitate the release of the PDMS mold from the surface [129]. The whole process takes about 20 to 60 minutes depending on the thickness of the photoresist. The master can after that be used repeatedly to create microfluidic chips by providing a mold in which a synthetic polymer, often polydimethylsiloxane (PDMS), can be solidified (see Figure 18). In this work I used two different microfluidic devices, one which allows fast switching (within seconds) but is not well suited for long time experiments and another one with can be used for very long experiments but has slower switching dynamics.

3.2.1 Microfluidic chip to investigate signaling dynamics in yeast

To investigate and control the signaling dynamics of the HOG cascade, I used a microfluidic device in which the media of yeast cells can be altered within less than a second. This device has already proven its usefulness in probing dynamical aspects of the HOG pathway [38]. The device has a simple Y-shaped structure and the cells can be fixed to the glass slide of the device with the sugar binding protein concanavalin A which also binds to glass as well as to carbohydrates on the cellular membrane. The device has two inlets (top of the Y, see Figure 19 for a schematic representation of the device) and one outlet and the media is flown through the device by applying a pressure to the input reservoirs with a microfluidic pressure controller (MFCS, Fluigent, Paris (France)). To change the input fluid that is felt by the cells, the fact that the flow in a microfluidic device is laminar is exploited. At larger scales, fluids exhibit a turbulent flow which shows a chaotic behavior and in which (fast) mixing occurs predominantly through convection. At the

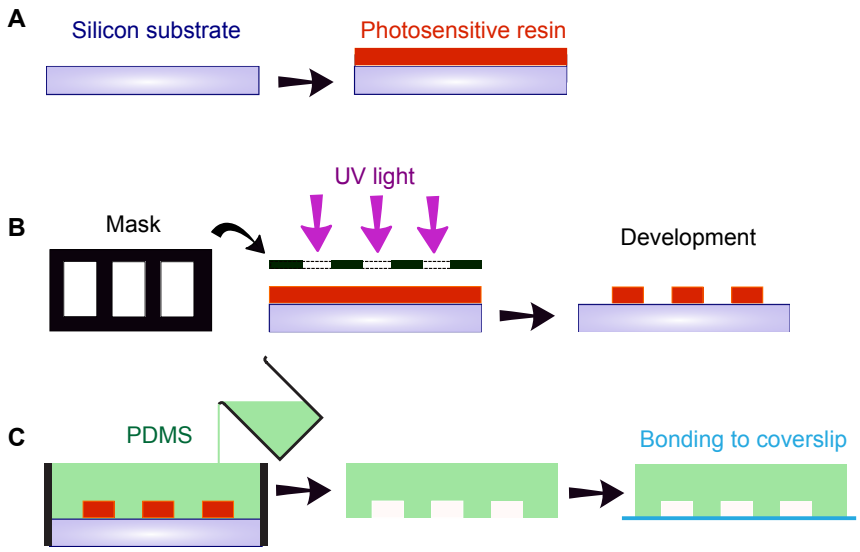


Figure 18: Microfabrication process. **(A)** A silicone wafer is covered with a photoresist. A regular thickness is achieved by rotating the wafer. The thickness of the photoresist layer is determined by its viscosity and by the rotational speed. **(B)** A pattern is transferred from a mask to the photoresist by UV light. After that the non-exposed part (negative photoresist) can be removed by a developing solution. **(C)** The pattern on the wafer can be used to construct microfluidic devices consisting of PDMS. Therefore fluid PDMS is applied to the master, polymerized and subsequently peeled of. After that the PDMS can be attached to a coverslip. (Modified illustration. Original by Agnès Miermont)

micro-scale, a fluid shows a completely different behavior in which the flow is laminar and (slow) mixing occurs only through diffusion. In the case of the microfluidic device presented here this means that the two fluids entering the device do only mix by diffusion and show a clear separation if the flow rate is high enough. The position of the line of separation can be altered by changing the pressure difference between the two input fluids, which makes it possible to rapidly (within less than a second) change the fluid which is flowing over cells close to the center of the device. A limitation of the Y-shaped device is that cells tend to loose their concanavalin A bonding to the glass slide after about 2 hours, which prevents the use of this device in long time experiments which are necessary if one wants to investigate gene expression.

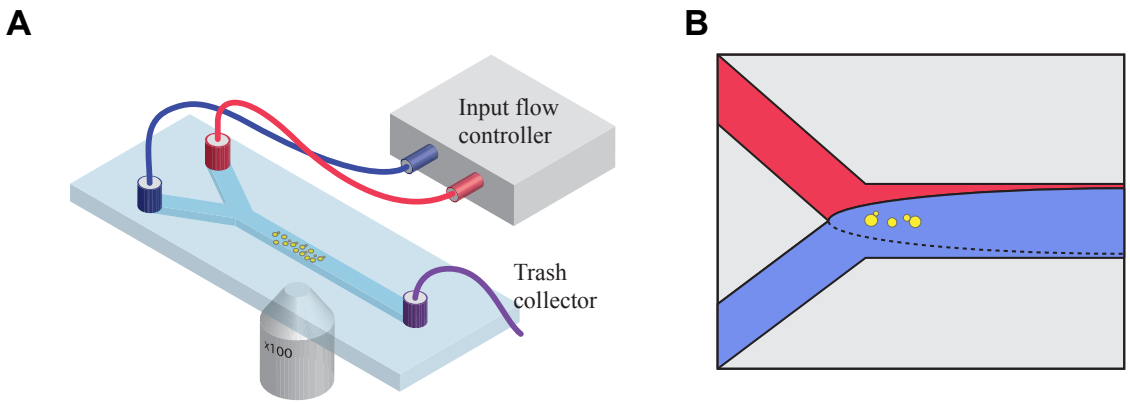


Figure 19: Y-shaped microfluidic device. **(A)** Schematic representation of the device. The flow is driven by the input flow controller which regulates the pressures of the two input fluids. Cells are fixed to the glass slide by the lectin concanavalin A, so they are not washed away by the flow. **(B)** The separating line between the two input fluids can be altered by changing the pressure difference between the two input fluids. This allows the exchange of the media flowing over the cells within less than a second.

3.2.2 Microfluidic chip for long term observation in yeast

Changes in gene expression occur on a much longer time scale than changes in the signaling activity of a cell, which means that the Y shaped device is not well suited for the investigation of gene expression. Hence, I used a different device for the control of gene expression, which allows long time observation of yeast cells and has an H-shaped form (see Figure 20). For the fabrication of the device I used an original design by Gilles Charvin. In this device the flow is driven by a peristaltic pump through two flow channels which have a height of 100 μm. Imaging chambers with a height of 3.1 μm are placed in between these flow channels. Since the diameter of a haploid yeast cell is approximately 3-6 μm^2, cells in this chamber are sandwiched in between the glass slide and the elastic PDMS layer. In this constrained environment, cellular movement is limited, which facilitates cell-tracking and allows long time observation of the cells. Cells can be loaded in this device by applying a short small pressure, which lifts up the elastic PDMS layer, allowing the cells to enter the imaging chamber.

The fluid entering the device is selected by a two-way valve (LFA series, The Lee Company, Westbrook CT, USA) upstream of the device and media exchange from the flowing channel to

[2]http://http://bionumbers.hms.harvard.edu

the imaging chamber occurs by diffusion. Since the fluid takes some time to travel from the valve to the device, a switch of the valve does not lead to an instantaneous change of the media felt by the cells. In addition fluids mix in the tubes and the media exchange in the imaging chamber happens by diffusion. Overall this means that a step change at the valve level will lead to a delayed and gradual change at the cell level (see Figure 20B). On the other hand the device is well suited for long time experiments, because cells growing out of the imaging chamber are just washed away by the flow.

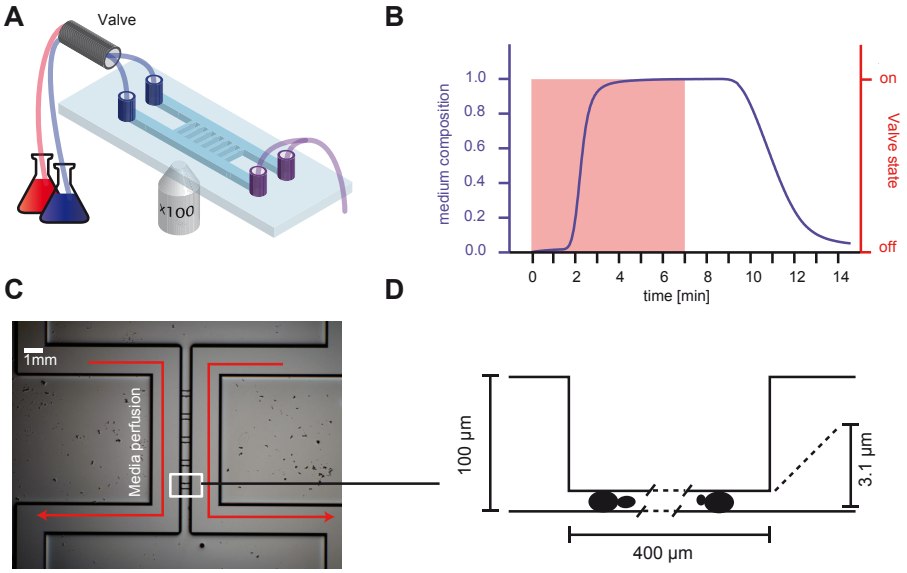


Figure 20: H-shaped microfluidic device. **(A)** Schematic representation of the device. The flow is created by a peristaltic pump placed downstream. It is possible to select between two different input fluids by switching a valve placed upstream. **(B)** Measurement of the switching characteristics using ink. A switch of the valve (red) does not lead to an immediate change at the microfluidic device level (blue curve), because of the dead volume in the connection tubes and because of mixing within the tubes **(C)** Close up of the microfluidic device. **(D)** Cells are trapped within a thin layer in between the cover slip and the PDMS. The media in this imaging chamber is exchanged by diffusion.

3.2.3 Manufacturing process

The master wafers for the two microfluidic devices were constructed with the standard soft lithography techniques described above. The photoresist I have used was SU-8 (MicroChem, Newton MA, USA) together with the processing protocol provided by the manufacturer. The microfluidic chips were constructed by casting PDMS (Sylgard 184, Dow Corning, Midland MI, USA) on the master wafer. The PDMS was prepared using a 10:1 ratio of the elastomer and the curing agent and subsequent removal of bubbles by vacuum degassing. The PDMS was cured at 65 °C overnight. After that the chip was cut out and bonded to a glass cover slip by plasma activation.

3.3 Yeast strains

For the control of the HOG signaling cascade I used a strain in which Hog1 is labeled with GFP and the nuclear histone Htb2 is labeled with mCherry (yPH15). The strain for the gene expression control experiments using the HOG cascade (yPH91) has the coding sequence of *STL1* replaced by yECitrine and in addition Hog1 tagged by mCherry. I constructed this strain with the help of Thierry Delaveau based on a strain provided by Megan McClean. In addition to gene expression control using the HOG system, I will show in Chapter 6 that the platform can be extended to use another gene inducible system, the MET3 promoter. The strain used for the gene expression control using the MET3 system (yPH113) carries a Venus fluorescent protein with a degradation tag controlled by the MET3 promoter. Furthermore, the strain has the budneck marker *MYO1* labeled with mCherry. All strains used here are listed in Table 3.1.

Strain	Genotype	Background	Reference
yPH15	HOG1::GFP-HIS3 HTB2::mCherry-URA3 MATa	S288C [130]	[40]
yPH91	pSTL1::yECitrine-HIS5 Hog1::mCherry-hph MATα	S288C [130]	[43] [3]
yPH113	pMET3::Venus-deg-ADE2 MYO1::mCherry-HIS5 MATa	W303	[131] [4]

Table 3.1: List of yeast strains.

[3] This strain is based on a strain that was kindly provided by Megan McClean.
[4] This strain was kindly provided by Gilles Charvin.

4

Towards controlling gene expression: control of signal transduction and computational investigations

4.1 Control of the Hog1 signaling cascade

My first step in developing a closed loop control platform for gene expression was the implementation of a feedback controller for the HOG pathway activity. At the time when I started this project such an *in vivo* control of a signaling pathway had not been demonstrated and showing that controlling the HOG cascade is feasible would be a first step in utilizing feedback control approaches to apply precise perturbations to dynamical biological systems. I will use the HOG pathway to drive gene expression and at the time of development of the HOG pathway controller I assumed that being able to control HOG signaling activity was a necessary step for the devel-

opment of an effective gene expression controller. In addition, the HOG cascade shows a significantly faster response time than the downstream gene expression process, which reduces the time for tuning and testing the control platform. Many methods which I applied to control the HOG pathway activity are similar to the methods required to control gene expression, so developing a controller for the signaling process helped to guide the development of the gene expression controller.

The idea was to investigate which activation profiles of Hog1 are achievable using a closed loop controller, in order to then tailor an approach to control gene expression which takes into account the constraints identified during the development of this first controller. In particular I was interested in the effectiveness of different approaches to encode gene expression profiles by modulating the HOG pathway activity. One option to encode gene expression would be to work with a constantly active HOG cascade and to vary its activation level in order to reach a desired gene expression behavior (amplitude modulation). Another option would be to repeatedly activate the pathway for short durations and to vary the frequency of these activations (frequency modulation). In the following I will test both approaches, to investigate which one is suited best for a control of gene expression.

4.1.1 Setup

The platform to control Hog1 signaling activity consists of a microfluidic device to activate the HOG pathway, a fluorescent microscope to observe Hog1 localization, and of a computer implementing image analysis and the controller. I used the Y-shaped microfluidic device presented in Section 3.2.1, which is able to switch between two different cellular media within less than a second, but cannot generate a mixture of the two inputs (input is digital). But we can exploit a feature of the HOG pathway to emulate continuous input. The HOG pathway is not able to faithfully follow a fast signal but instead acts as a low pass filter (cutoff frequency 4.6×10^{-3} Hz which amount to a typical time-scale of 220 seconds) [38]) and integrates the signal. We can use this feature to emulate a continuous input by switching between normal and sorbitol-enriched (1M) medium faster than the cutoff frequency of the HOG cascade. In practice I used a pulse width modulation with a time window of 10 seconds, which would for example encode a 0.2 M sorbitol intensity by flowing cells 2 seconds with 1 M sorbitol media and 8 seconds with normal media.

The yeast strain I used for this control has Hog1 fused to GFP and and the nuclear protein Htb2

fused to mCHerry (strain yPH15 from Table 3.1), so the activation state of the HOG cascade can be quantified via the nuclear colocalization of Hog1 (see Section 3.1.3). To identify cellular regions I used a thresholded image of Hog1::GFP, while a thresholded image of the Htb2::mCherry channel was used to identify nuclear regions. In this work I did not identify single cells, but instead controlled the signaling activity of the few (1-6) cells in the field of view, wherefore cells are not tracked over time.

I used a PI controller (see Section 2.3.2) to control the HOG cascade, which due to its simplicity is easy to implement and shows a good performance for most control applications. The derivative term of the PID controller was omitted due to the high measurement noise of the observation of Hog1 activation, which otherwise would have a negative influence on the robustness of the controller. Because I am considering a tracking problem here, in which the target changes over time, only the recent past errors are relevant for control. Therefore, I integrated the error only in the interval $[t - \Delta, t]$, where Δ is 100 seconds. The controller was tuned manually using a trial and error approach. I found a good trade-off between response time and stability of the controller by setting the proportional term k_P to 2 and the integral term k_I to 1.5, good trade-off meaning that the control system does not show excessive overshoot or oscillatory behavior, while the response time is reasonably fast.

4.1.2 Results

To test the two control strategies discussed above (amplitude vs. frequency modulation of the signal transduction output) I designed two control objectives. The first was to maintain the nuclear localization at a constant level, which was 20% higher than the nominal value in unstressed cells (this value was chosen such that it poses a significant increase of the colocalization, while being lower than the value for the maximal activation of the pathway). This control target simulates an amplitude modulation, in which the HOG cascade is active for a prolonged period. The second control objective was designed to simulate a frequency modulation control of gene expression by activating the HOG cascade successively for a short duration. In practice I used trapeze like motifs, which resemble the natural response of the HOG cascade to a pulse of osmolarity. The amplitude of the trapeze was again 20% higher than the nominal value and the increase and decrease times were 2 minutes respectively, while the plateau duration lasted 3 minutes.

The results of these two control experiments in Figure 21 clearly show that the control is effective. For example if we consider the step experiment at time 2 minutes, the controller first applies the maximal osmotic input which results in an increase of the Hog1 nuclear localization after a time delay of about 1-2 minutes. The system overshoots slightly, so the input is reduced

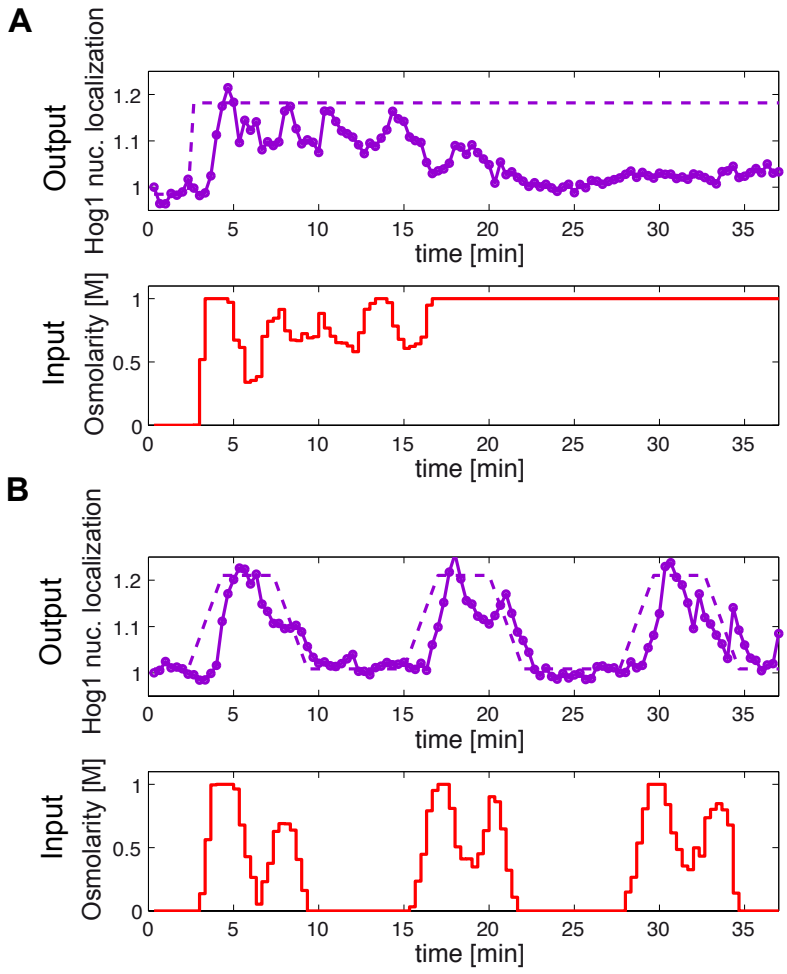


Figure 21: Experimental results for the control of Hog1 nuclear localization. **(A)** Controlling the Hog1 localization to a constant value 20% higher than the normal value does only work for short times due the feedback mechanisms activated by Hog1. **(B)** When activating the HOG cascade in a repeated manner, with relaxation periods in between the activations, successive trapeze signals can be achieved.

by the controller. After that follow oscillations just below the target value for about 10 minutes. After these 10 minutes the controller applies the maximal stress, while the activation of the HOG pathway slowly falls back to its initial value. This behavior can be explained by the feedback mechanisms of the HOG cascade (see Section 2.1.3). Prolonged activation of Hog1 leads to the production of glycerol, so the cells adapt to the osmotic stress and the HOG cascade is inactivated and the cells become insensitive to high osmolarity. It has been shown that prolonged activation of the HOG cascade is possible by a ramp input [70], but since the maximal stress that can be applied is bounded this is not an option. This also means that the amplitude modulation based control strategy is not feasible. In addition the activation state of the HOG cascade is almost always lower than the target value. This can also be explained by the osmo-adaptation of the cells, which renders the osmo-stress less and less effective with time.

In contrast the controller succeeds in reproducing the trapeze like motifs, which encode a frequency modulated Hog1 signal. Here the 6 minute relaxation time between two pulses seems to be enough time to reset the pathway. This might be because under iso- and hypo-osmotic conditions the glycerol channel Fps1 opens, so the produced glycerol can leak out of the cell. Again there is a 1-2 minute time delay of the Hog1 response, which could be easily circumvented by shifting the target function back in time. Taken together these results indicate that the frequency encoding strategy is more suited to our problem of controlling gene expression than an amplitude modulation of Hog1 activity.

4.2 Computational investigations

A control system requires rigorous testing and tuning before it will function properly, a procedure that can be both time consuming and costly. In particular testing control approaches for biological systems can be very time consuming. For example a problem with the gene expression control experiments that I will present in the following chapter was the long duration of these experiments. Gene expression changes take place over long time, so each experiment lasted at least 15 hours while the "memory" of the cells eliminated the possibility of restarting an experiment in case anything went wrong. This rendered the development and debugging process of the real-time controller quite slow-moving. In practice I was able to run at most one gene expression control experiment in *S. cerevisiae* per day. Another factor was that gene expression is a noisy process, so developing a control approach necessitates several experimental replica to eliminate stochastic ef-

fects. On the other hand it is possible to simulate control systems *in silico*, which greatly facilitates the development process. It is straightforward to evaluate the effectiveness of different controller setups with a computer, given that a satisfying model of the process to be controlled is at hand.

After developing a platform for the feedback control of the HOG pathway activity, the next step would be to implement such a platform for the control of gene expression. Because of the difficulties mentioned above, I first investigated the applicability of different control approaches *in silico*, before controlling gene expression in live cells. In the following section I present a two-layered approach for gene expression control using the HOG cascade, which exploits the fact that the signaling cascade and its gene expression response operate on different time-scales and can be treated independently. This approach guided the development of the gene expression control platform presented later, even though it was not adopted wholesale.

4.2.1 Model development

To simulate the behavior of a gene expression controller we first need a model of the underlying process. We can then use this model to evaluate the controller. In addition the controller is based on a model predictive control (MPC) approach, which relies on a model of the system, to predict the effect of possible control strategies. Several models of the HOG cascade have been proposed (see Section 2.1.3), but most of these are not suited for controlling gene expression because of two problems. Firstly, most of the models developed for the HOG pathway describe the signaling cascade, while omitting the gene expression response. Secondly, the complex structure of most HOG pathway models renders state estimation infeasible. In a biological system, only a limited numbers of variables can be measured simultaneously *in vivo*, so for controlling the system the model needs to be simple enough to recover the full state from the limited variables that can be observed. For this reason I developed a simple switched linear model, which relates an osmotic input profile to the resulting gene expression response. While the HOG signaling cascade responds within about 3 minutes to an osmotic step shock, the gene expression response is much slower [66]. This time-separation can be used to partition the process into two connected modules, a signal transduction module and a gene expression module. This modular structure can be used to simplify the control problem. While the signal transduction module provides the input of the gene expression module, the two parts can otherwise be treated independently. The

structure of the signal transduction module is as follows:

$$\text{if } o_{ext}(t) \leq o_{int}(t): \quad \text{(iso- or hypoosmotic conditions)}$$

$$\dot{h}(t) = -\gamma_g h(t) \tag{4.1}$$

$$\dot{o}_{int}(t) = k_o h(t) - (\gamma_o + \gamma'_o) o_{int}(t) \tag{4.2}$$

$$\text{if } o_{ext}(t) > o_{int}(t): \quad \text{(hyperosmotic conditions)}$$

$$\dot{h}(t) = k_g(o_{ext}(t) - o_{int}(t)) - \gamma_g h(t) \tag{4.3}$$

$$\dot{o}_{int}(t) = k_o h(t) - \gamma_o o_{int}(t) \tag{4.4}$$

In this model the only input is the relative external osmolarity $o_{ext}(t)$, while $o_{int}(t)$ describes the relative intracellular osmolarity whose variations result from glycerol synthesis and degradation/export, and $h(t)$ denotes the activation of the HOG cascade corresponding to the nuclear localization of the Hog1 protein. The relative osmolarity is defined here as the difference between current and steady state value in normal media. In hyperosmotic conditions, the Hog1 activation is assumed to be proportional to the intensity of the hyperosmotic stress felt by the cell $o_{ext}(t) - o_{int}(t)$, while in hypoosmotic conditions Hog1 is not activated. The inactivation of Hog1 is proportional to the amount of active Hog1. Active Hog1 has a positive influence on the production of glycerol, which is modeled by a proportional term $k_o h(t)$. The glycerol channel Fps1 is only open under hypoosmotic conditions, wherefore the term $\gamma'_o o_{int}(t)$, which describes glycerol diffusion through Fps1 is only present in these conditions. In addition I assume a proportional degradation of glycerol $\gamma_o o_{int}(t)$, which is always present. Gene expression is modeled by a simple reaction-based model depicted in Figure 22. It describes protein and mRNA levels, while assuming an exponential degradation of both entities. mRNA production is proportional to the amount of active Hog1, while protein synthesis is proportional to the mRNA concentration. The model can either be interpreted as a stochastic model using the Gillespie algorithm [132, 133], or as a deterministic model with the following ODEs:

$$\dot{r}(t) = k_r h(t) - \gamma_r r(t) \tag{4.5}$$

$$\dot{p}(t) = k_p r(t) - \gamma_p p(t) \tag{4.6}$$

This is a standard way of modeling gene expression that has been applied in a similar manner various times [134].

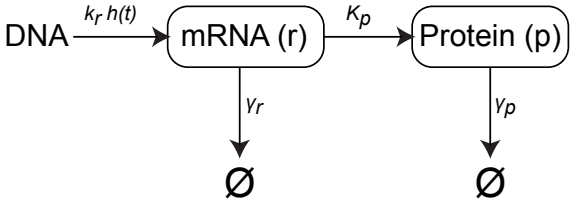


Figure 22: Gene expression model. Production of mRNA is proportional to Hog1 activation ($h(t)$), while protein prodution is proportional to the mRNA level. Both mRNA and protein are degraded exponentially.

The parameters of the signal transduction model have been determined by minimizing the mean square deviation between model output and experimental data consisting of the localization response of Hog1 to different osmotic inputs (see Figure 23). For fitting, the global optimization algorithm CMAES (see section 2.4) was used. Parameter values are listed in Table 4.1. For the gene expression model realistic parameter values have been used, because at the time of the construction of this model, I did not have any gene expression data available. The values of the degradation parameters correspond to mRNA and protein half-lifes of 5 and 20 minutes respectively.

k_o	γ_o	γ'_o	k_h	γ_h	$o_i(0)$	$h(0)$
0.005	0	9.05	48	1.21	0	0.019

k_h	γ_r	k_p	γ_p	$r(0)$	$p(0)$
0.1386	0.1386	0.3466	0.0347	0	0

Table 4.1: Parameter values and initial conditions for the signal transduction (top) and for the gene expression (bottom) models.

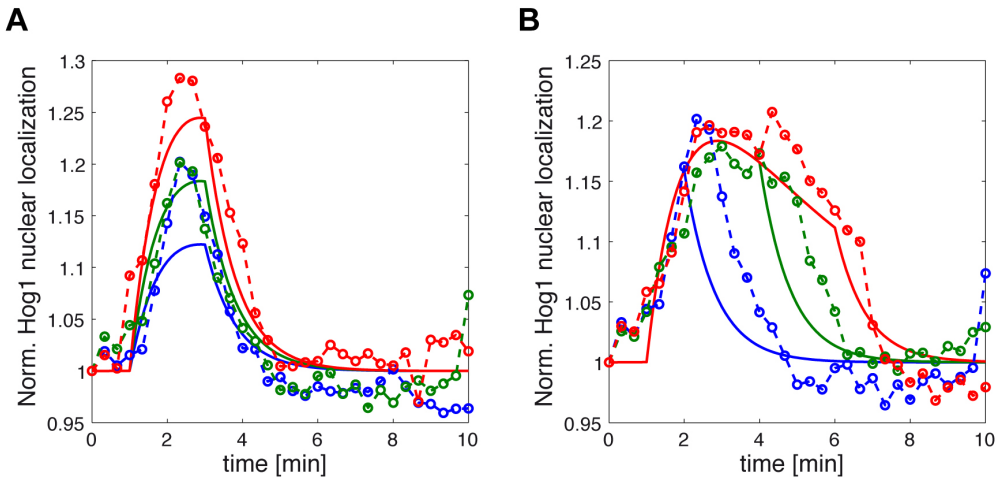


Figure 23: Cell response to different hyper-osmotic stresses and simulation of the HOG pathway model. **(A)** Stresses of different intensities. Blue, green, and red plots correspond to 0.4, 0.6, or 0.8M stress applied during 2 minutes. **(B)** Stresses of different durations. Blue, green, and red plots correspond to 0.6M stress applied during 2, 4, or 6 minutes. Dashed and solid lines represent experimental data and model predictions, respectively. All stresses started at time 1 minute.

4.2.2 Control strategy

Two variables of the model described above can be observed experimentally by fluorescent labels: the localization of Hog1 and the gene expression output of the system, which are the outputs of the signal transduction and gene expression modules. This can be exploited to divide the control problem into two smaller ones, which can be solved independently (see Figure 24). In a first step we can search for a Hog1 profile that when applied to the gene expression system leads to the desired gene expression profile. Thus, we regard the output of the signaling cascade as the control input to the gene expression system. Afterwards the controller searches for an osmolarity profile which implements the Hog1 profile found in step 1. This backstepping approach allows one to treat the two subsystems independently, which is advantageous because they operate on different time scales. Thus different time horizons for the control prediction and different sampling times can be used for the two subsystems.

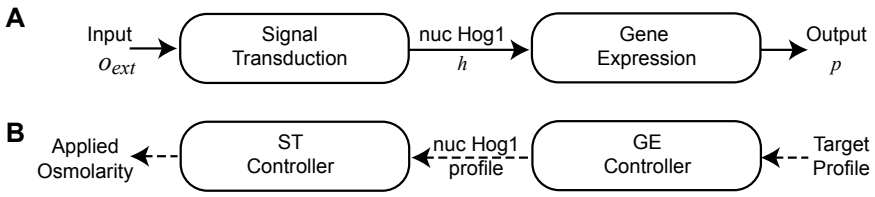


Figure 24: **(A)** The osmotic stress response system can be split into two subsystems: A relatively fast signal transduction system and a relatively slow gene expression subsystem. The output of the signal transduction system, the Hog1 nuclear localization acts as the input for the gene expression system. **(B)** Separation of the control problem into two simper ones based on the separation of the model.

Constraints and parametrization

Following the results of the control experiments of the HOG cascade in Section 4.1, which showed that a prolonged activation of the HOG pathway is not achievable, I constrained Hog1 profiles to be repeated trapeze motifs, which have a maximal duration and have to be separated by a minimal time. A Hog1 profile is described by the number of trapezoidal motifs n, a vector of trapeze durations d_{on} and a vector d_{off} describing the off durations (see Figure 25). In addition there are fixed times for the increase and decrease, for the minimal time between two activations, and for the duration of the plateau of 2, 2, 5 and 1 minute respectively.

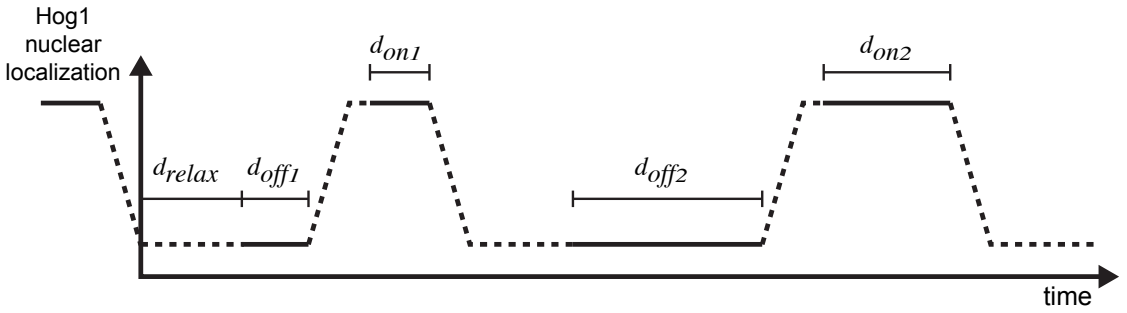


Figure 25: Hog1 profile constraints. A Hog1 profile is made up by reoccurring trapeze motifs, parametrized by a vector of plateau durations d_{on} and a vector of off-times d_{off}. The rise- and fall-off-times of the trapeze are fixed. In addition there is a minimal trapeze duration and a minimal relaxation time. All times that are not variable are indicated as dotted lines.

The input of the signal transduction module is given by a piecewise constant osmolarity profile $o_e(t) = u_i$ if $\tau_i \leq t < \tau_{i+1}$ and $u_i \in [0, 1]$ with the discretization $(\tau_1, \ldots, \tau_m)$ (see Figure 26).

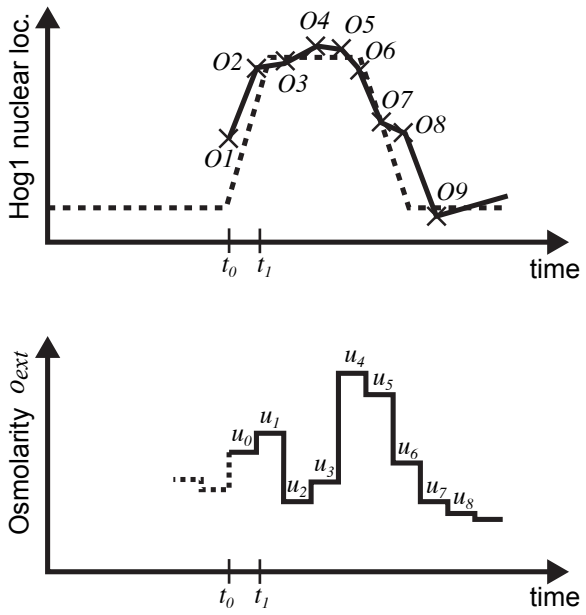


Figure 26: Realization of a trapezoidal Hog1 profile by a piecewise constant osmotic input. Starting from the trapezoidal Hog1 motif found during the first step of the control process (top, dotted line), the model predictive controller searches for a piecewise constant osmolarity profile which will implement the found Hog1 profile. In this example the value of the control u_0 is found based on the observation $O1$. After that a new observation is made and a control value for the new interval is chosen.

Algorithm

Using the backstepping strategy, the controller starts by identifying a Hog1 profile, which when implemented will lead to the desired gene expression profile. This Hog1 profile is found by minimizing the mean squared deviation (MSD) between the gene expression model output and the target protein profile within the next time window (in this case 100 minutes). The parameters optimized here are the number of trapezes n, their durations d_{on} and the off-times d_{off}. For the parameter search the global optimizer CMAES is used. In practice one parameter search is done

for each value of $n = 1, 2, \ldots, k$ and the profile with the best fit is selected. This is implemented in the matlab function `[`n`,`d_{on}`,`d_{off}`] = searchHogProfile(`t`,`s_{ge}`,`*target*`)` with the current time t, the current state of the gene expression model s_{ge} and the protein profile *target*.

Once a Hog1 profile has been found, the controller searches for the osmolarity profile that results in the desired Hog1 profile. The control input is again found by a parameter search using CMAES, but in this case the time horizon is shorter than the one for the gene expression profile (here 2 minutes). The osmolarity parameters $u_1, \ldots u_m$ are found by the function `u = searchOsmProfile(`t`,`s_{st}`,`t_p`,`n`,`d_{on}`,`d_{off}`)` with the current time t, the state of the signal transduction model s_{st} and t_p, n, d_{on}, d_{dur} defining the Hog1 profile with the time point of the first pulse t_p.

The separation of the control problem into two smaller ones greatly facilitates the computation of a control strategy, because it allows for different prediction horizons for the two modules. For the slow gene expression response the d_{on} and d_{off} parameters are identified for a time horizon of 100 minutes, which is necessary, because gene expression changes occur slowly. Identifying all u_t parameters of the osmotic input that would implement the complete Hog1 profile found in the fist step would be computationally infeasible. In addition it is not necessary to compute the osmolarity for such a long time period, because HOG signaling reacts fast so that the influence of the control can be observed rapidly.

4.2.3 Results

To test the suitability of this control approach I applied it to drive the same model which is used by the MPC. Of course this is not a fair evaluation of the quality of the controller because the same model is used for control and evaluation, but it gives an insight whether the proposed control strategy might be able to drive gene expression in live cells. The signal transduction module was simulated using the ODE interpretation, while the gene expression module was either interpreted as a deterministic ODE system or as a stochastic system using the Gillespie algorithm. A reason for using only deterministic simulations for the signal transduction module is that most of the variability of the system stems from gene expression, while the signaling cascade shows a more or less deterministic behavior. Two control targets were selected, first to maintain protein concentration at a given level and second to make the protein level follow a sinusoidal signal. The results are shown in Figure 27 and indicate the feasibility of a Hog1 pulse modulated control ap-

proach. Even in case of the stochastic interpretation of the gene expression module, which shows significant levels of noise, the protein level stays within admissible bounds around the target.

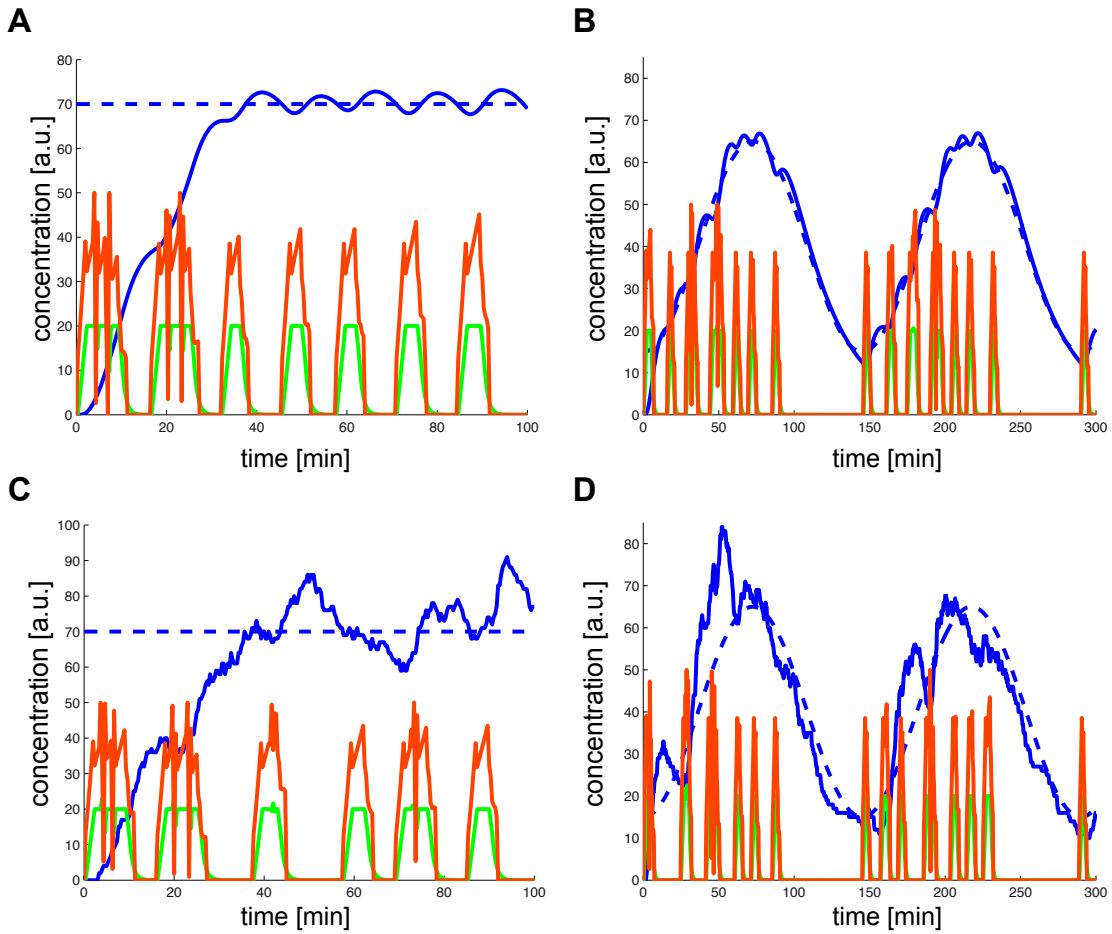


Figure 27: Gene expression control simulation. The dotted blue line represents the target protein profile, the solid blue line the observed protein level, the green line the nuclear level of Hog1 and the red line the osmotic input (x50). The plots show deterministic control simulations with a constant (**A**) or time sinusoidal (**B**) protein target and stochastic control simulations for the same target profiles (**C** and **D**).

4.3 Conclusion

As a first step towards the feedback control of gene expression, I have presented here a control platform for the HOG signaling cascade, which to the best of my knowledge was the first attempt to control a signaling cascade in real time and *in vivo* at the time of publication (January 2011, [40]). Controlling the Hog1 localization to fixed target values is not possible for extended times due to the feedback mechanisms triggered by active Hog1, while repeated activations of the cascade, with relaxation periods in between the activations, are feasible. This indicates that a control of gene expression via the HOG cascade will only work if it does not require the constant activation of Hog1. This result suggests the use of a frequency encoded Hog1 signal to achieve a certain gene expression profile, in contrast to an amplitude encoded signal in which Hog1 needs to be constantly active to some degree.

In the second part of this chapter I presented a computational investigation of a gene expression control strategy which uses a frequency encoded Hog1 signal. In this approach the Hog1 signals are constrained to be trapeze like motifs, with a minimal and maximal plateau time and a minimal relaxation time between consecutive Hog1 activations. Exploiting the different time-scales of the gene expression and signal transduction processes, the controller separates the control problem in two smaller ones following a backstepping approach. In a first step, given a desired protein profile, the controller finds a Hog1 profile. In a second step the osmolarity profile leading to the desired Hog1 profile is identified. Using an MPC strategy, this controller works well in *in silico* simulations even in the presence of realistic levels of noise. In the next chapter I will focus on the implementation of a similar control strategy *in vivo*. The results of the computational investigation have been published [135].

5

A platform for the control of gene expression

5.1 Towards a gene expression control platform

In the previous chapter I presented a platform to control HOG signaling activity as well as computational investigations of how this platform could be extended to control gene expression. In this chapter I will describe the development of such a gene expression control platform. First experimental tests quickly revealed that a simple adoption of the computational method proposed in the previous chapter was not feasible due to experimental constrains. Firstly, the method proposed assumes that the osmolarity of the cellular media can be changed almost instantaneous and in a continuous manner. This can be emulated with the Y-shaped device which I used for the control of the HOG cascade, via a pulse width modulation. But first gene expression experiments using this device showed that it is not suited for experiments lasting longer than about 2 hours, because after that time cells tend to detach from the glass slide (mainly because of cell-division). This was the reason for using the H-shaped device described in Section 3.2.2 which is well suited for long time experiments. On the other hand the H-shaped device has a slow switching time

(about 3 minutes) which forbids the emulation of a continuous osmolarity signal by a pulse width modulation strategy as I used it for the control of the HOG cascade. A second problem was the assumption that the localization of Hog1 could be measured over the whole experiment. The high imaging frequency required to capture the dynamics of Hog1 in combination with the long time scales on which gene expression changes take place makes it unfeasible to monitor the HOG pathway activity over the whole course of a gene expression control experiment. For this reason, the controller presented here does not take into account the HOG pathway activity, but relies on gene expression measurements as the sole output of the system.

On the positive side, the MPC approach, in particular the frequency encoding of gene expression to limit natural adaptation, can be adopted with only slight modifications. Because the only output of the system which is measured is the gene expression response, state estimation might not be possible for the model I presented previously, wherefore I use a simplified model here. The two-layered control approach, which treats the signaling cascade and the gene expression response independently has not been used, because the output of the HOG cascade cannot be observed for prolonged time periods.

The goal of developing this control platform is to demonstrate the ability to control gene expression *in vivo* with high accuracy using the HOG signaling cascade for both time-constant and time-varying targets. Because of the natural feedback mechanisms involved in the HOG cascade, this would demonstrate that control of a biological system is feasible even in the face of strong perturbations. Fixed control targets are interesting because they allow one to study the effects of well-controlled, steady perturbations of a cellular protein level. Time varying perturbations on the other hand can be used to probe the dynamics of a system. In particular they could be used to decipher the functioning of gene networks [37, 128].

5.2 Experimental setup and controller development

This section describes the setup of the gene expression control platform. The microfluidic device used here is the H-shaped device described in section 3.2.2.

5.2.1 Yeast strain

In the yeast strain I used for this work (yPH91, see Table 3.1) the *STL1* gene has been replaced by the fluorescent marker yECitrine. *STL1* codes for a proton glycerol symporter located in the cell membrane which is involved in the active uptake of glycerol from the media. Deleting *STL1* causes poor growth on glycerol based media. Despite the fact that importing glycerol should help cells to survive stress, osmosensitivity was only observed in combination with other deletions of genes involved in osmotic adaptation [136]. *STL1* is one of the genes most strongly activated by osmotic stress [34], without responding to the general stress response transcriptional factors Msn2 and Msn4 [35]. An activation by the general stress response could pose a problem in control experiments, because this could lead to unmeant activations of gene expression by other possible stresses. The strong activation in response to osmotic stress together with the independence from the general stress response make *STL1* a perfect candidate for a gene control using the HOG cascade. In addition Hog1 has been fused to the fluorescent protein mCherry in the strain I use to observe the localization of Hog1. Even though this information cannot be used for gene expression control due to the bleaching problems mentioned above, it has been used for different preliminary investigations. This was in particular useful to get an idea of the Hog1 activation profiles for different shock durations in the H-shaped microfluidic device.

5.2.2 Image analysis and fluorescent quantification

For the detection and tracking of cells I have used the circular Hough transform described in Section 3.1.1 and the tracking method described in Section 3.1.2. Levels of pSTL1-yECitrine fluorescence have been quantified as described in Section 3.1.4.

5.2.3 Modeling

Because I will use a MPC strategy to control gene expression, a model of the system that relates the osmotic input to the observed gene expression output is required. Since the only output we can observe is the gene expression response given by the fluorescence of yECitrine, it might be a problem to reliably estimate the state of the model developed for the computational investigation in Section 4.2.1. For this reason I developed a simple two-dimensional ODE model which has

the following form:

$$\dot{x}_1(t) = u(t-\tau) - g_1 x_1(t) \tag{5.1}$$

$$\dot{x}_2(t) = k_2 x_1(t) - g_2 \frac{x_2(t)}{K + x_2(t)} \tag{5.2}$$

In this model x_1 denotes the recent osmotic stress felt by the cell and x_2 the protein fluorescence level. The function $u(t-\tau)$ describes the osmotic input which is shifted by τ minutes, thereby aggregating time delays caused by signal transduction, gene expression and protein synthesis, and folding. In principle the delays for gene expression and protein synthesis should appear in equation 5.2, however in this specific case it is mathematically equivalent and computationally more efficient to group all delays in a unique term in the input. The osmolarity function u is based on a piecewise linearization of the microfluidic switch dynamics which have been measured using ink (see Figure 20B). The recent osmotic stress x_1 is modeled by a term integrating the osmotic stress ($u(t-\tau)$) and a linear decay term ($-g_1 x_1(t)$), which lets the influence of past osmotic stresses diminish exponentially with time. The fluorescence level increases linearly with x_1 with rate k_2. The degradation term of the fluorescence is modeled by a saturating Michaelis-Menten kinetic. This is a non-standard way of modeling protein decay, which is usually assumed to be exponential due to dilution caused by exponential cellular growth. In my experiments I observed that in response to repeated osmotic shocks the cellular fluorescence increased linearly over a time of more than 10 hours (see Figure 28). This observation is not in agreement with an exponential decay term using realistic protein turnover rates if we assume a quasi constant production rate in the condition of repeated osmotic shocks. Because it has been shown that the protein degradation machinery can saturate *in vitro* [137], I assumed a saturating degradation term. On the other hand experiments showed that the pSTL1-yECitrine expression often reacts stronger to osmotic inputs towards later times during an experiment, which suggests that long time memory effects might be responsible for the observed dynamics.

The parameters of the model have been determined by fitting the model to different characterization experiments using the global optimization tool CMAES. I exposed the cells to two different types of input profiles to determine the dynamics of the Hog1 activated *STL1* expression. In the first type I sent a single osmotic shock lasting between 5 and 8 minutes every 4 hours, in order to get an idea of the effect of single pulses and to observe the protein degradation dynamics in

this system. The second type of characterization experiments consisted of osmotic shocks sent repeatedly every 30 minutes, to observe the maximal production capability of the system. Based on these observations, I set the time delay τ to 20 minutes and the protein level corresponding to degradation half-saturation K to 750. I chose the value of K, such that degradation does not saturate in the single shock experiments. The model shows a fair but not perfect fit to the data (see Figure 28), which is already good given that this model describes a complex signaling pathway including its gene expression response with just two variables. The parameter values of the model are listed in Table 5.1.

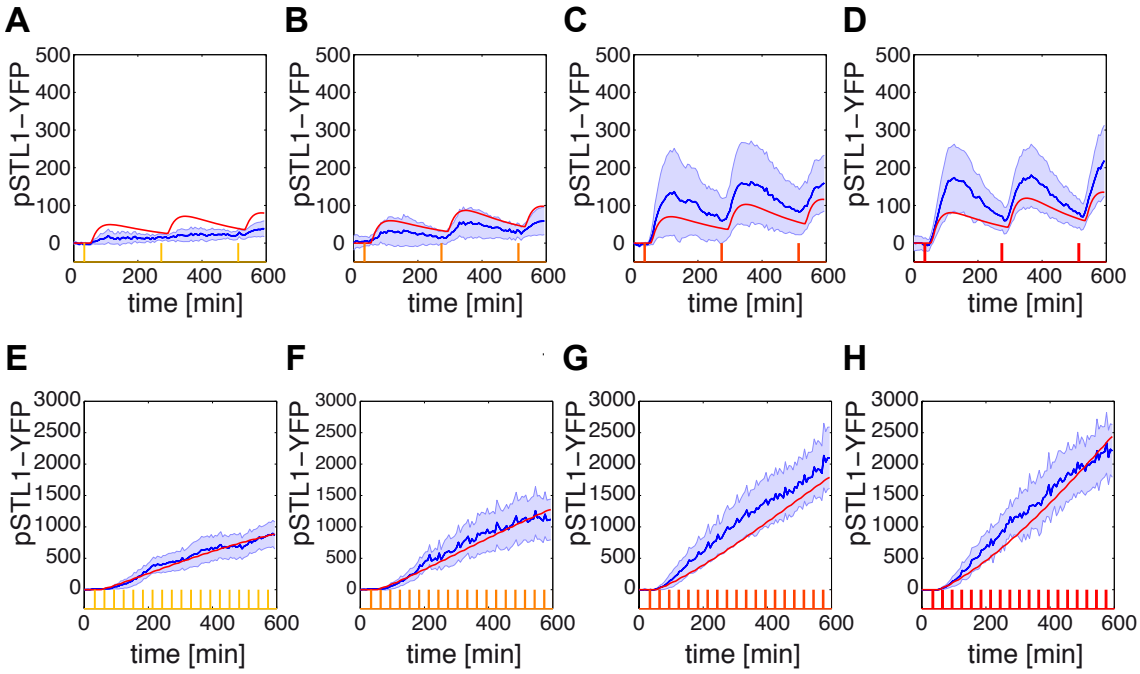


Figure 28: Calibration gene expression data and model fit. **(A-D)** Isolated osmotic shocks lasting 5,6,7 and 8 minutes applied every 4 hours. **(E-H)** Osmotic shocks lasting 5,6,7 and 8 minutes applied every 30 minutes. Fluorescence of yECitrine under control of the *STL1* promoter was quantified. Mean and standard deviation are shown as blue lines and blue envelopes. The model was fitted to the mean. Model simulations are shown in red. Note that the fluorescence intensities reached in repeated or isolated shock experiments are significantly different

g_1	g_2	τ	k_2	K
4.02×10^{-3}	37.5	20	0.851	750

Table 5.1: List of parameters.

5.2.4 State estimation

For state estimation I used an extended Kalman filter (EKF) (see Section 2.3.3) with a measurement noise (R) of 2500 and the covariance matrix of the process noise (Q) diag(0.37,925). The value of R has been chosen by estimating the variance of the observations from experimental data. The ratio of the diagonal values in Q has been determined by the steady state values of x_1 and x_2. The actual values of Q have been determined by ensuring consistency between the innovation residuals (the difference between model prediction and observation) and the variance of the innovation residuals, as estimated by the Kalman filter.

5.2.5 Control strategy

The control strategy I use for gene expression control is similar to the one presented in the previous chapter, with the exception that no measurement of Hog1 localization is taken into account. Similar to the previous strategy, activations of Hog1 are limited in both duration and frequency to prevent the accumulation of glycerol and the resulting adaptation of the cell. In contrast to the previous control experiments, in which a continuous osmolarity signal could be emulated using fast switches and by exploiting the low pass filtering capabilities of the cells (see Section 4.1.1), the H-shaped device only allows a relatively slow switch between normal and sorbitol-enriched media. To prevent adaptation mechanisms I limited the shock durations to values ranging from 5 to 8 minutes, and ensured that consecutive pulses are separated by at least 20 minutes. The 5 minute lower bound ensures that the cells actually feel the stress, because shorter stresses might be diluted by the time they have traveled from the valve to the imaging chamber. These input constraints are one reason for using MPC, which allows easy integration of constraints. A schematic representation of the control platform and the MPC strategy is shown in Figure 29.

The fluorescence of yECitrine is observed every 6 minutes and the state of the system is estimated using the EKF. Using the current state of the model, the controller searches then for admissible osmolarity profiles minimizing the mean squared deviation between the model output and

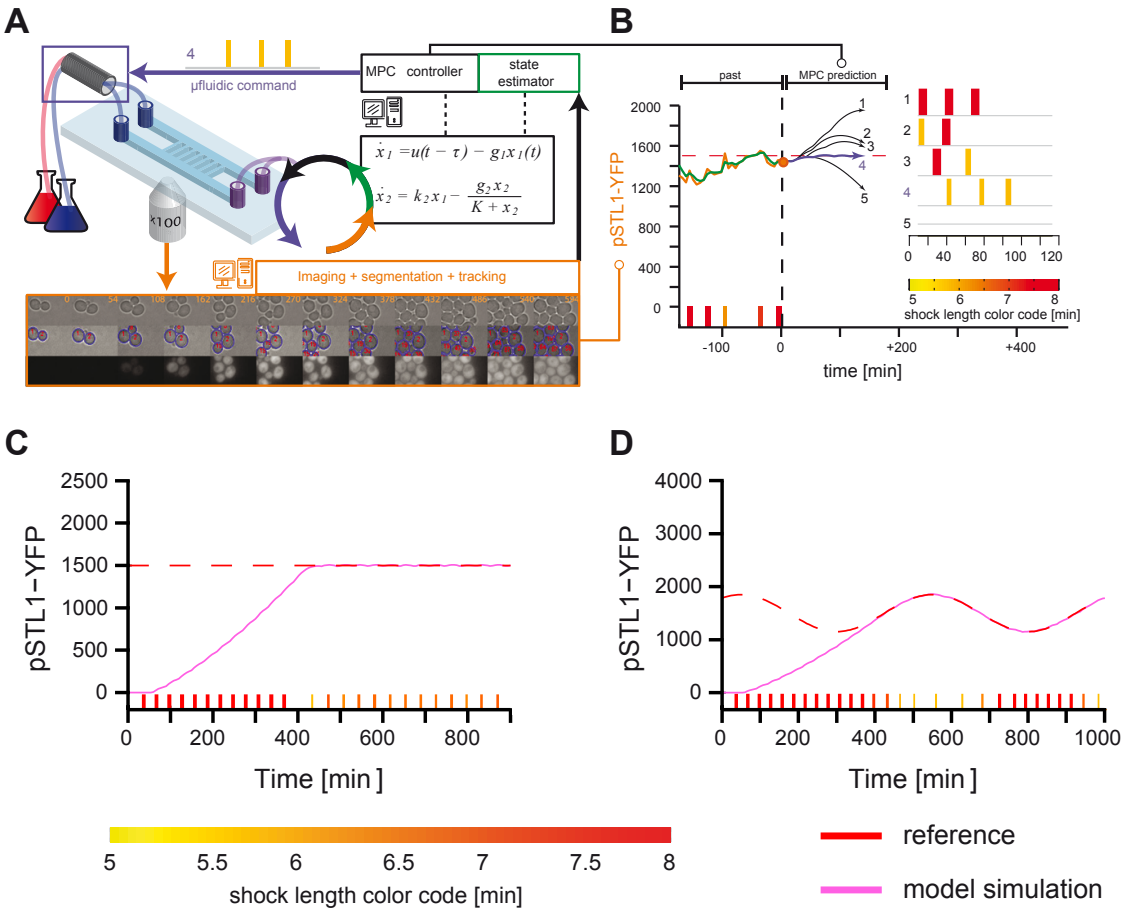


Figure 29: Gene expression control platform. **(A)** Schematic representation of the control platform. Yeast cells are placed in a microfluidic device, which allows the alternation between normal and sorbitol enriched media. A computer does automated imaging, segmentation and cell-tracking using a Hough transform. The expression of the fluorescent reporter yECitrine is sent to a state estimator. A model predictive controller computes the osmolarity profile which is applied to the cells using a mathematical model of pSTL1-yECitrine induction. **(B)** MPC search strategy. At the current time-point (orange circle), the state of the system is estimated (green) and the MPC searches for the best input to apply to the system. The best input strategy is identified by minimizing the distance of the model output (black curves) to the target protein profile (dashed red line) for different input profiles. An input profile is specified by the number, starting times and durations of osmotic pulses. A measurement is done every 6 minutes. **(C-D)** Computer simulation of a control experiment for a time-constant **(C)** and time-varying **(D)** target profile. The controller is driving the same model as used for control. The reference target profile is shown as a dashed red line and the result of the simulation as a pink line. The results indicate that the controller implementation is effective.

the target protein profile over a 120 minutes time horizon. This problem is recast into a parameter search problem in which an osmolarity profile is characterized by the corresponding stress start times and durations. In practice a single optimization problem is solved using CMAES for each number of stresses considered (1-3), while taking into account the predicted performance of not applying any stress during the 120 minute time horizon. Because image analysis and computation of the best control strategy may take up to 3 minutes, the input is not applied immediately when the image is taken but 3 minutes after this observation. An input is only applied in case the controller found that a shock should be applied within the next 6 minutes (before the next control computation result is available). If nothing is to be applied, computation of the best control strategy is repeated for the next observation.

To test whether this control strategy yields an effective feedback controller, I tested it *in silico* in the same manner as in the computational investigation in the previous chapter. I again used the MPC algorithm to drive the same model of the systems that is also used to predict its behavior. In this case the system state is estimated by the EKF based on the observation of only the gene expression output of the model x_2. The results in Figure 29C-D show an almost perfect control, indicating that the proposed MPC strategy is effective.

5.3 Experimental results

Using the control platform described above, I controlled the cellular fluorescence of yECitrine in either populations of all the cells within the microscopic field of view, or alternatively in one single cell arbitrarily chosen at the beginning of the experiment. In principle the control of a population of cells should be easier than single cell control because stochastic effects average out. But single cell control should be more precise at the single cell level, because the controller can directly react to the output of the controlled cell. All experiments lasted at least 15 hours, starting with only few cells in the beginning, while ending with 100-300 cells in the field of view.

5.3.1 Population control experiments

I chose two different types of targets: (i) to maintain the average fluorescence level at a given constant value (set-point experiment) or (ii) to make it follow a time-varying target profile (tracking

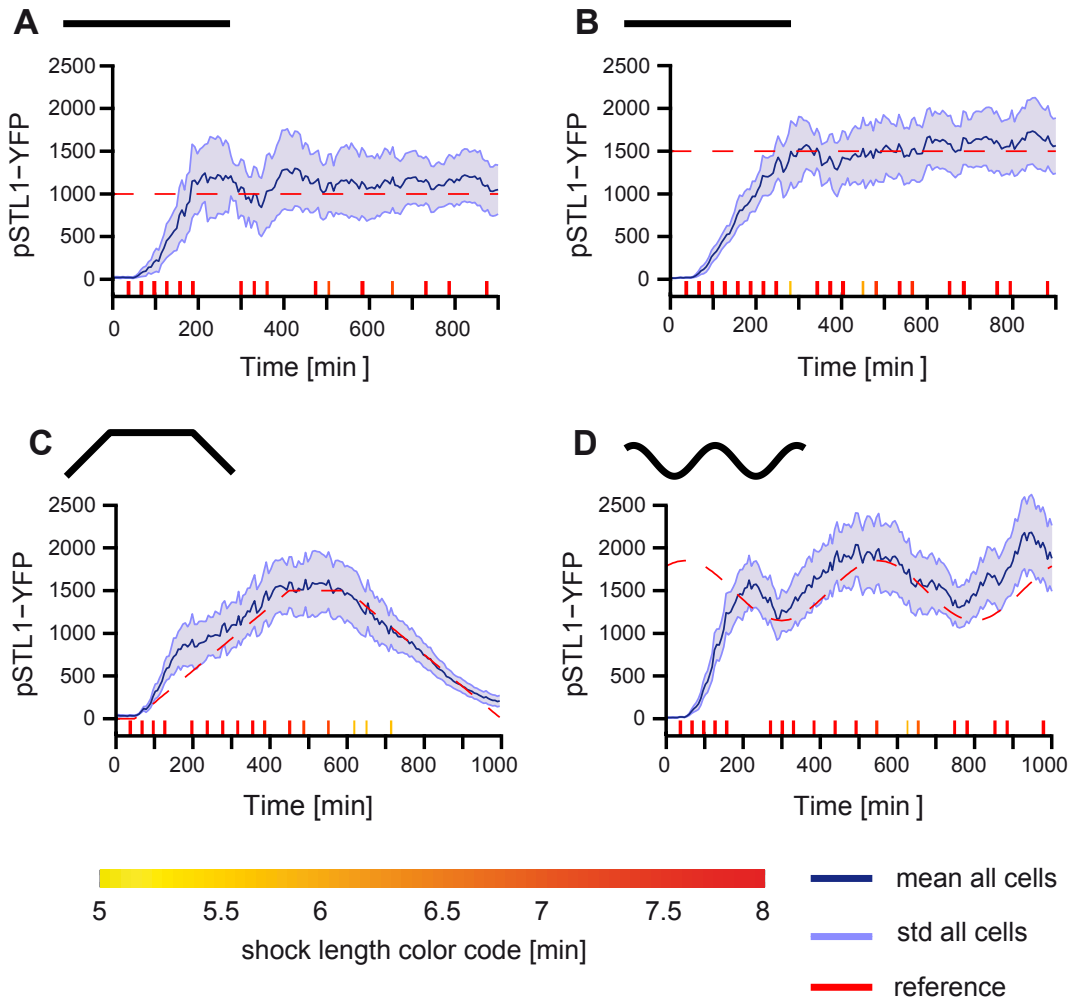


Figure 30: Control of gene expression at the population level. For all plots the target protein level is indicated as a red dotted line and the observed fluorescence of pSTL1-yECitrine is indicated as a blue line with light blue envelopes indicating the standard deviation. The osmotic input profile is indicated at the bottom of each plot and is color coded. (**A** and **B**) Set point control experiments with target values 1000 and 1500 f.u. (**C** and **D**) Tracking control experiments. The profile in (**C**) has a trapezoidal shape with a maximum of 1500 f.u. and the profile shown in (**D**) shows a sinusoidal shape.

experiment). It can be seen in Figure 30 that the controller works properly for gene expression control at the population level. For the set-point experiments a certain time is required to reach the target value, after which the average fluorescence stays within reasonable tight bounds around the reference. The lowest target value I tested is 200 fluorescence units (f.u.) and the highest is 2000 f.u. (see Figure 31), showing that the controller performs well within a 10-fold range, even though there is an overshoot for the low target value (200 f.u.). To evaluate the effectiveness of the feedback loop, I also tested an open loop strategy, in which the control input is computed before the experiment based only on the model. Figure 32 shows that this approach does not work, the feedback loop is required for precise control in this system. The deviation between the model simulation and the actual behavior of the system indicates the effectiveness of the feedback loop, but also highlights that the model does not capture very well the dynamics of the system, which is not surprising for a simple two-dimensional model.

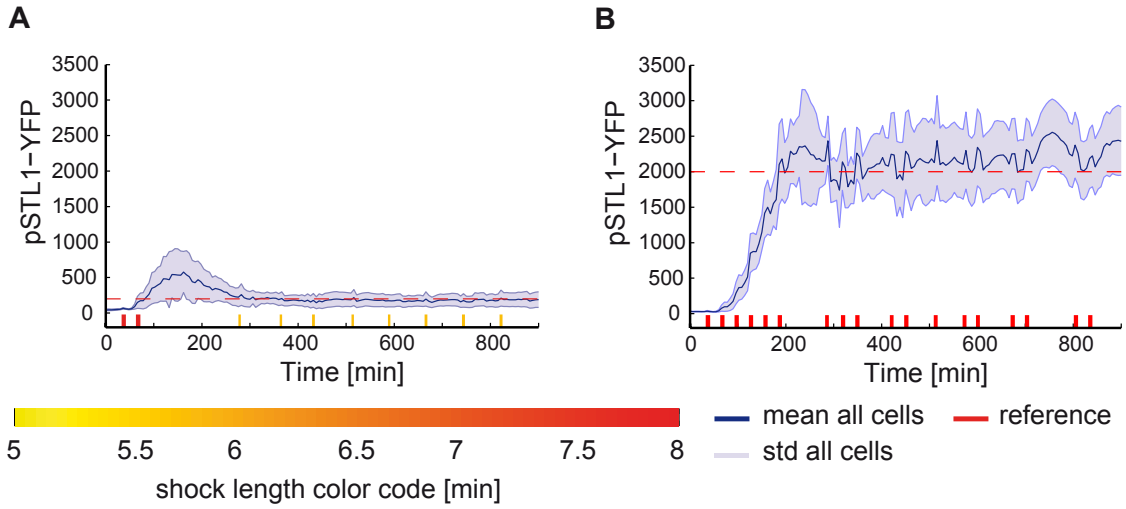


Figure 31: Gene expression control range. To determine the upper and lower limits in which the controller can drive gene expression, set point control experiments with high and low target values were conducted. **(A)** Set point population control with a target value of 200 f.u. The control shows a significant overshoot in the beginning, but after ∼300 minutes, the observed fluorescence follows the target faithfully. Lower control objectives are not possible because the target would be too close to the background fluorescence of the cells which is around 30 f.u. **(B)** Set point control to test the upper limit of the control range using a target value of 2000 f.u. The control works but shows significant levels of noise.

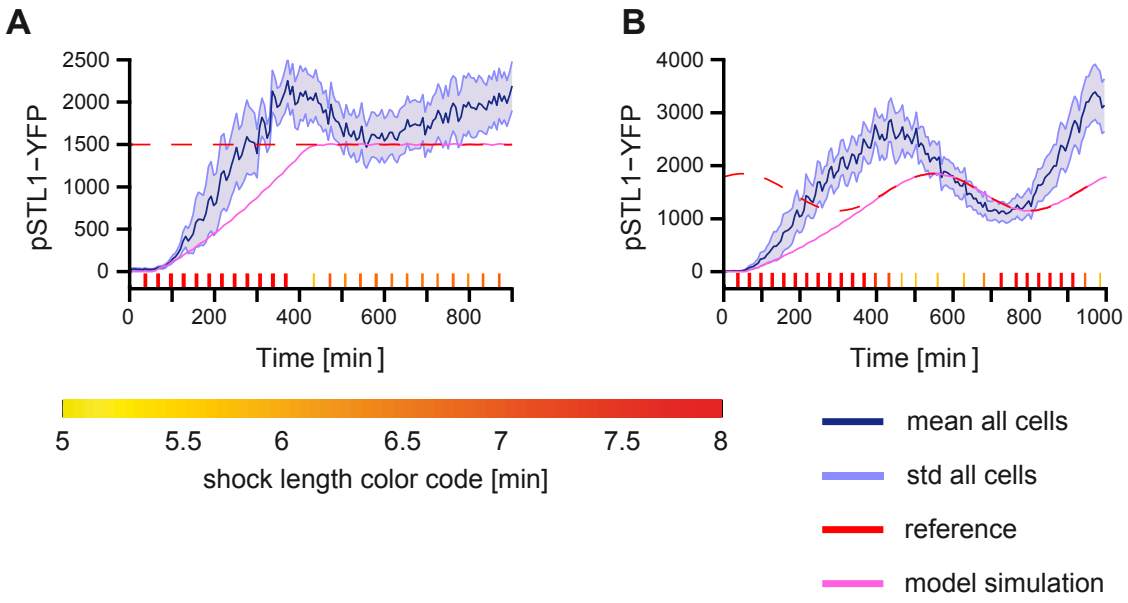


Figure 32: Open loop control experiments. To test if a feedback controller is really necessary to control gene expression, the control was computed beforehand based on a simulation of the model and the found control strategy was applied to a population of cells. The model simulation (pink line) shows an almost perfect control, indicating that the model predictive control approach is functional. When applying the control strategy to live cells significant deviations between model simulation and the observed fluorescence were found. **(A)** Open loop control of a fixed target at 1500 f.u. The observed fluorescence does not follow the target and does not even converge to a fixed value. **(B)** For the open loop control using a sinusoidal target, the observed fluorescence shows oscillations with the right frequency, but the amplitude is not captured properly and the oscillation is out of phase. These results indicate that closed loop control is required for a precise gene expression control.

For the tracking experiments I chose a trapeze function with the plateau at 1500 f.u. and a sinusoidal wave also oscillating around 1500 f.u. In particular sinusoidal protein expression profiles will be an important tool to decipher the dynamic behavior of gene networks. The results of the tracking experiments in Figure 30 indicate that the controller is able to make gene expression follow a varying target. Of course the admissible target functions are constrained by the maximal induction and degradation rates of this system. For the trapeze, the increasing slope is less steep than what the system is able to deliver. Overall these results indicate that the controller is functional.

5.3.2 Single cell control experiments

In addition to controlling cell populations, I was interested whether control was possible at the single cell level. Many biological questions, for example whether cells express genes in a continuous or in a burst-like manner, can only be answered by looking at individual cells rather than population averages [138]. Being able to control gene expression in single cells would help to generate precise perturbations, which would help to investigate single cell dynamics.

For the control of single cells I manually chose one cell at the beginning of the experiment, which is then followed by the tracker. In the rare case that the cell is lost, the experiment is aborted. The target profiles were the same as for the population control. As shown in Figure 33, the platform also performs well for single cell control, even though single cell control is *a priori* more difficult than population control, because gene expression noise is not averaged out. Note that sometimes the controlled cell behaves significantly differently from the rest of the population (see for example Figure 33A). Comparing control inputs for different experiments with the same target value showed that the applied osmolarity profile varies considerably (see Figure 34), suggesting that closed loop control was necessary for the control of single cells.

One question that remains to be investigated is whether controlling a single cell gives better results compared to a cell in a population control experiment. We can define the mean square deviation (MSD) of a trace, either of a single cell or of a population mean as its average squared difference from the target value.

$$d_{msd} = \frac{1}{n}\sum_{i=1}^{n}(x_i - r_i)^2 \tag{5.3}$$

with the trace $X = (x_1, \ldots, x_n)$ and the target function $R = (r_1, \ldots, r_n)$. If we look at the MSD of either the fluorescence of the controlled cells from single cell control experiments or the MSD of the mean fluorescence in a population control experiment, population control performs clearly better (see Figure 35). But this comparison is not fair because noise in the population control experiment is averaged over the population. Instead we can compare the MSD of the single cell control to the average of the MSDs for the single cells in the population control experiments. Note the difference between the MSD of the population mean and the mean of the single cell

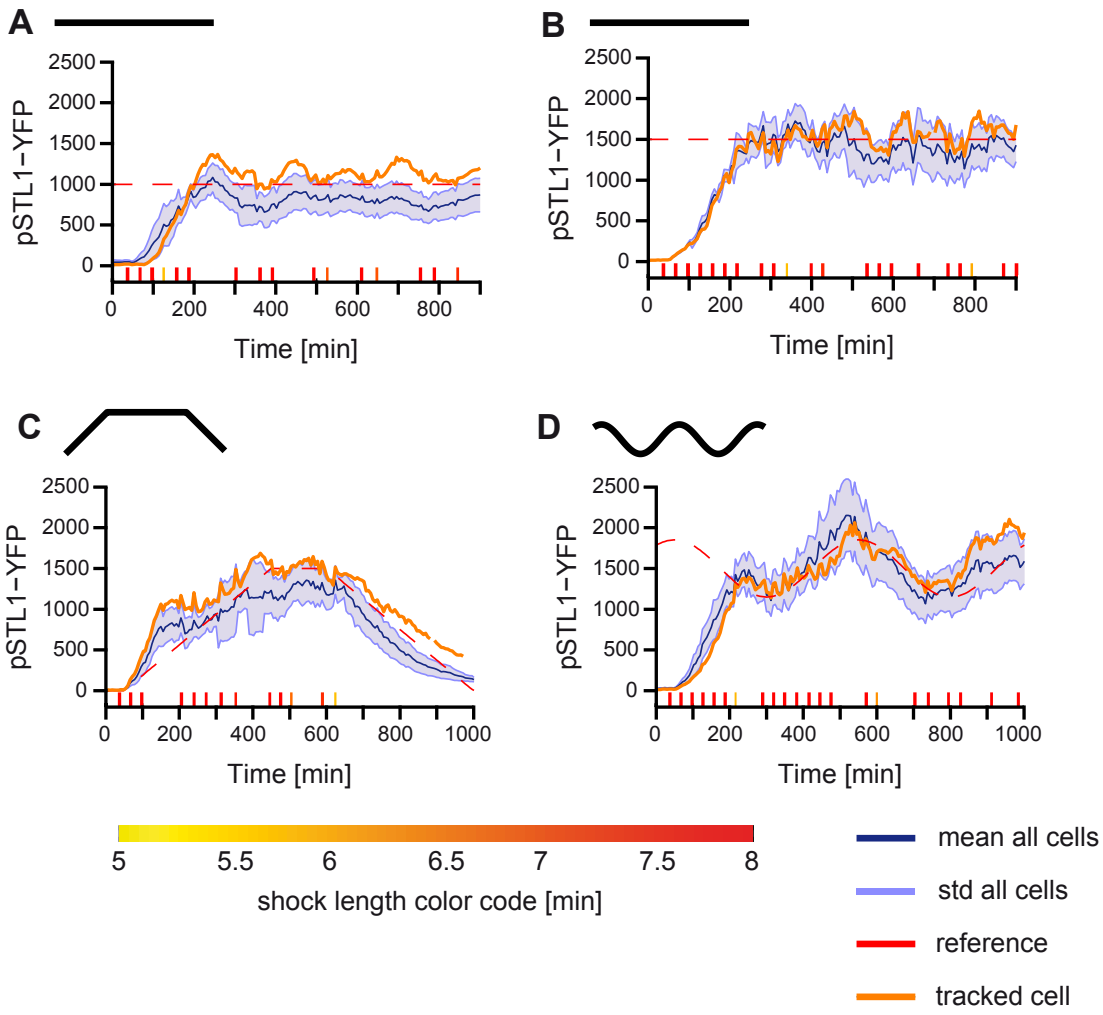


Figure 33: Control of gene expression at the single cell level. A single cell was arbitrarily chosen at the beginning of the experiment, and its fluorescence (orange line) was used to feed the controller. The average fluorescence and standard deviation of the population are shown as a blue line and as a blue envelope respectively. (**A** and **B**) Set point control experiments with target values 1000 and 1500 f.u. The population mean also follows the target but with less accuracy than the controlled cell. In some experiments (see **A**) the population mean behaves markedly different from the single cell. (**C** and **D**) Tracking control experiments. In **C** the control target is a trapeze with a plateau at 1500 f.u., while the target in **D** has a sinusoidal shape.

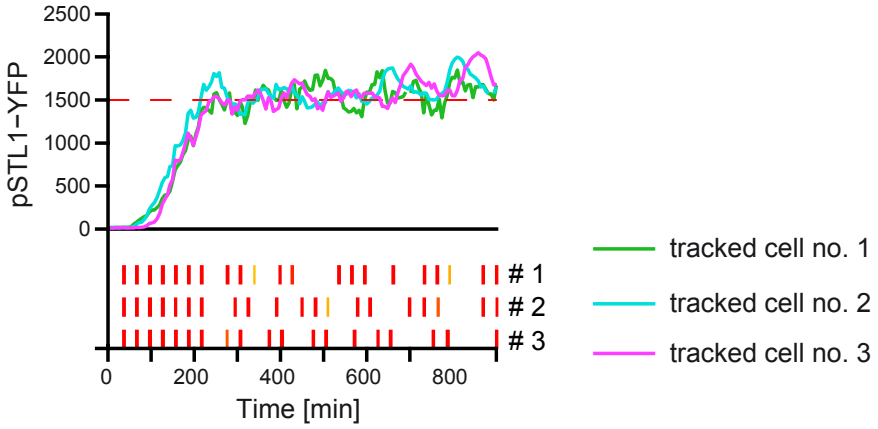


Figure 34: Repetitions of single cell control experiments. The control inputs for different single cell control experiments vary, even though the target value is the same (in this case 1500 f.u.). This highlights the need for closed loop control, because cellular variability requires different inputs to make different cells follow the target profile faithfully.

MSDs. In practice I omitted the data for the first 250 minutes for the set-point and sine-wave experiments, because of the time required to reach the target value. Also only cells which have been successfully tracked for at least 5 (set-point) or 7 (tracking) hours were considered. Figure 35 shows histograms of the MSDs of single cells in a population control experiment and compares them to the MSD of single cell control runs. The histograms indicate that the control performance of single control is better than the control performance for an average cell in a population control experiment (compare the orange bars (MSD of single cell control) to the black bars (average MSD of population control) in Figure 35). To have a quantitative answer, we can then use a statistical test to determine whether the distribution of the MSDs for single cell controls equals the distribution of the single cell MSDs in the population control experiments. Hence the null assumption (H_0) is that MSDs in single cell and population control experiments follow the same distribution. The alternative assumption (H_1) is that the MSDs are statistically smaller in the single cell control. Because of the unknown distribution of MSDs, a non-parametric test should be used to asses whether the two distributions are the same. A standard test whether two samples of observations have smaller or larger values than the other is the Wilcoxon-Mann-Whitney (WMW) test [139], which assumes the same distribution of the two samples under the null hypothesis. Because it is expected that the variance of the MSDs is smaller for the single cell control,

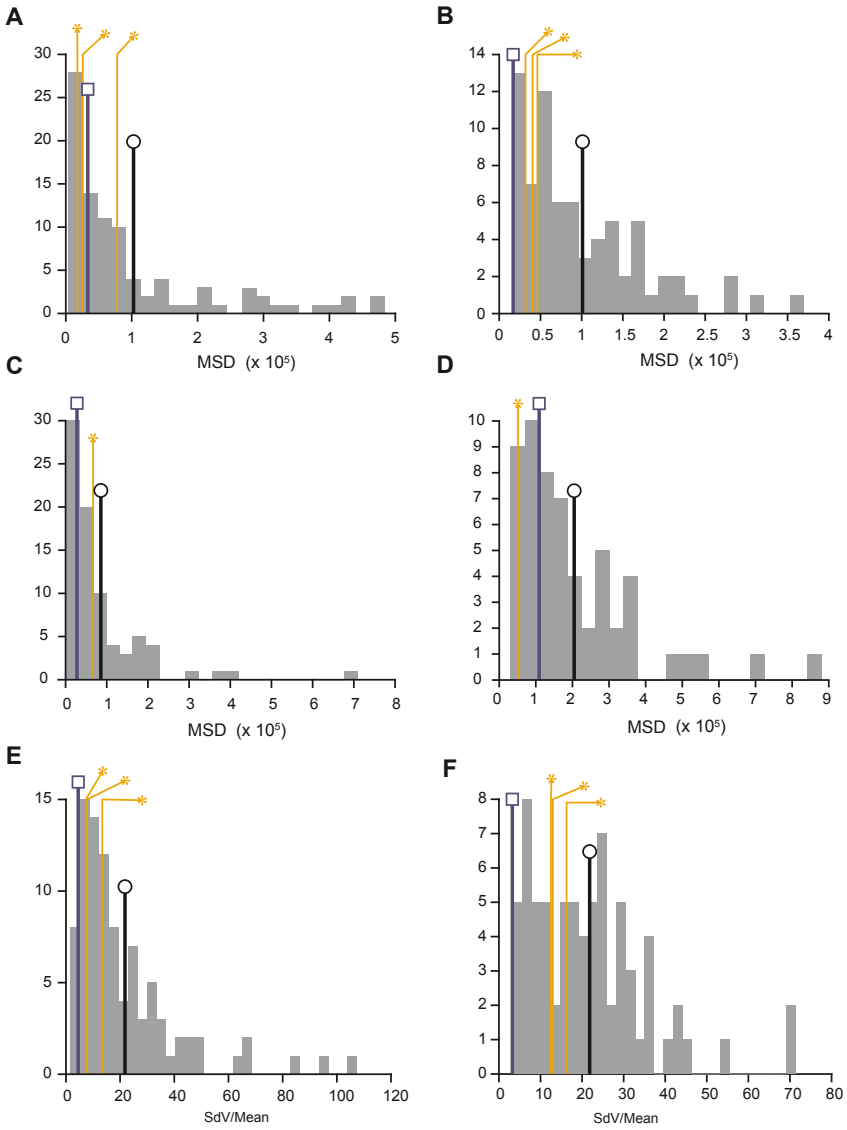


Figure 35: Comparison of control quality and noise level between single-cell and population control experiments. **(A-D)** Histograms showing the distributions of the single-cell MSDs in population control experiments. The MDS of the mean is indicated as a blue bar with a square and the mean of the single cell MDSs is indicated by the black bar with a circle. MSDs of single cell control experiments are shown as orange bars (*). **(A)** Target control at 1000 f.u. **(B)** Target control at 1500 f.u. **(C)** Trapeze control. **(D)** Sine wave control. (**E** and **F**) Histograms of the noise levels of the single cell fluorescence in the 1000 **(E)** and 1500 **(F)** set-point control experiments. The noise levels of the mean of the population are indicated as blue bars (with square), means of single cell noise levels as black bars (with circle) and noise levels for single cell control are indicated as orange bars (*).

Experiment	N1	N2	Ù	P value
MSD, T = 1000 f.u. (Figure 35 A)	92	3	171	0.15
MSD, T = 1500 f.u. (Figure 35 B)	73	3	172	1.7×10^{-7}
Noise level, T = 1000 f.u. (Figure 35 E)	92	3	202	3.7×10^{-3}
Noise level, T = 1500 f.u. (Figure 35 F)	73	3	144	8.5×10^{-3}

Table 5.2: Computation of the Ù statistic and its associates P value for MDSs and noise levels of the two set-point experiments using the FP test.

the WMW test does not apply to this case. Instead I used the Fligner-Policello (FP) test [140], which assumes the two samples to have the same median and to be symmetric, but does not rely on other assumptions about the distributions. When applying this test to the set-point experiments, it gives a highly significant result for the target of 1500 f.u. (p-value $1.7\cdot10^{-7}$), while the result for the target of 1000 f.u. is not statistically significant. The single cell control data available for the tracking experiments did not suffice to calculate statistics.

Another interesting question is whether single cell control reduces gene expression noise. For the set-point experiments, we can define the gene expression noise of a cell as the fluorescence standard deviation over the fluorescence mean. This value can be used analogical to the MSD to compare the distributions of single cell control and population control using the FP test. Results showed that for both targets the noise level was significantly reduced in case of single cell control (see Table 5.2). For histograms of the noise level see Figure 35E-F.

5.4 Effect of osmotic shock on growth rate

It is well known that hyperosmotic stress inhibits progression of the cell-cycle [141–143]. One mechanism is the phosphorylation of the cyclin-dependent kinase inhibitor[1] *Sic1* by active Hog1 which leads to an arrest in G1 phase [144]. In addition there is a mechanism leading to G2 phase arrest which is also Hog1 dependent [145]. When I checked the doubling time of the cells in the control experiments I have presented so far, this value was in between 100 and 250 minutes, which is slow compared to a normal doubling time of about 90 minutes. This cell-cycle slow

[1]Cyclin-dependent kinases are a family or protein kinases with important roles in the regulation of the cell-cycle.

down and the variations in growth rate can be explained by the osmotic input that is applied to the cells. Figure 36 shows the number of cells within the microscopic field for two different experiments, a trapeze control (Figure 36A) and a control to a fixed target of 1500 f.u. (Figure 36B). For the trapeze control experiment, strong osmotic stresses are applied during the increase and plateau period, while there is lower osmotic input for the decrease of the trapeze. After 800 minutes the whole field of view is filled with cells, so that no further increase of the cell number can be observed. For the fixed-target control experiment (Figure 36B), strong osmotic stresses are applied during the whole experiment to maintain the target value of 1500 f.u. This leads to a slower growth rate compared to the trapeze control experiment (~100 cells at 900 minutes for the fixed control compared to ~300 cells for the trapeze control).

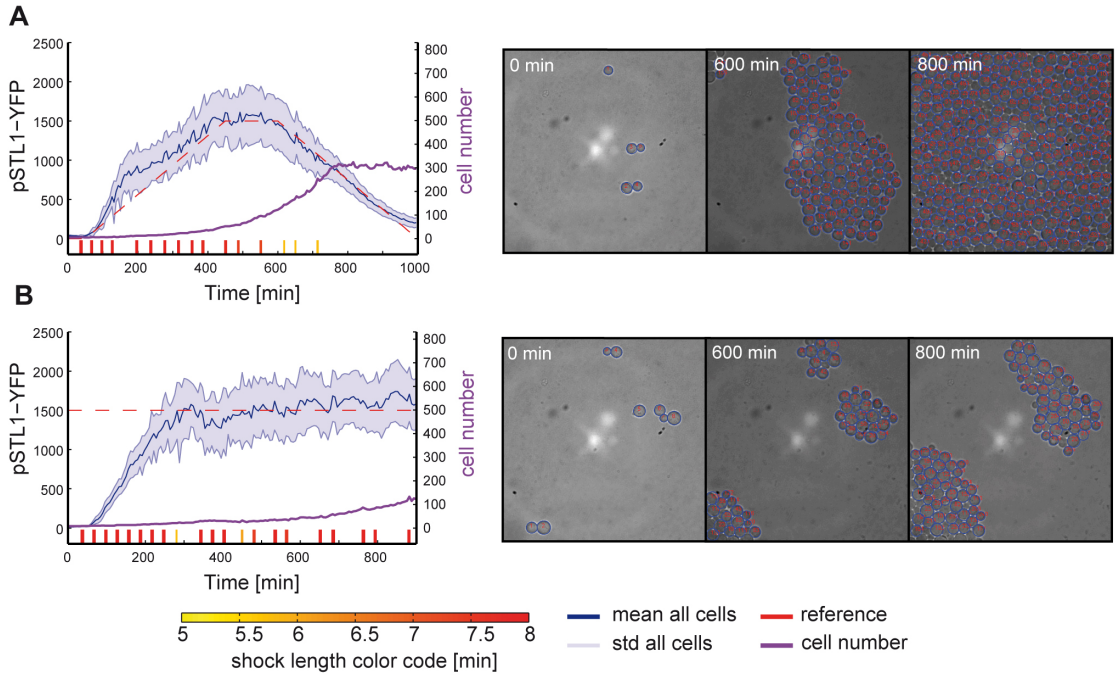


Figure 36: Cell growth during control experiments. **(A)** Number of cells (violet) within the microscopic field of view for a trapeze control experiment. The field of view is filled with cells after about 800 minutes. The maximal number of cells in the field of view is about 300. **(B)** A strong osmotic input is required to reach and maintain the target value 1500 f.u. in this control experiment. The strong osmotic input led to a slowdown of the cell-cycle compared to the trapeze control shown in (A).

The protein concentration inside the cells is constantly diluted by cellular growth, which has an impact on the observed degradation rate of a protein. The cellular growth rate in the control experiments I presented here is not necessarily constant over time, so that the protein dilution is also not constant over time. This should have a negative effect on the quality of the model predictions and thereby also on control performance. One solution would be to explicitly take the growth rate into account in the model.

5.5 Conclusions

In this chapter I presented a control platform, able to control gene expression in real time at both population and single cell levels. Even though the control quality might seem poor when compared to technical control systems, it has to be taken into account that I control a noisy biological system which could not be controlled before in live cells. In addition it would not be feasible to control gene expression without a closed-loop control setup, as illustrated by the failure of the open-loop control experiments (see Figure 32). There are mainly two reasons for the necessity of closed loop control for the precise control of gen expression: Firstly, gene expression shows a high degree of extrinsic noise, meaning that different cells react differently to the same input (compare Figure 33A and Figure 34). Together with the stochastic nature of the gene expression process (intrinsic noise), this means that even with a "perfect" model of the system it would not be possible to achieve high precision control, at least if the number of the controlled cells is low, so that stochastic effects are not canceled out by a large population. A second reason for the necessity of closed loop control is our inability to model gene expression processes at high accuracy, which is due to the complex and not yet completely understood dynamic behavior of biological systems. On the other hand we can correct for these prediction errors by constantly updating our control strategy based on observations of the system. Here I showed that in such a closed-loop control setting a good control quality can be achieved even though the model predicting the behavior of the system is very coarse compared to the complexity of the underlying processes. This means that the closed loop approach does not rely on a detailed model of the system, which also is illustrated by the fact that feedback mechanisms of the HOG cascade have not been included in the model. Note that a quantitative model suitable for control which includes these feedback mechanisms might not have been feasible, because of the various biological mechanisms involved (see Section 2.1.3). This does not mean that control could not be improved by incorporating cur-

rently un-modeled dynamics into the model. For example, it might be advantageous to include the cellular growth rate into the model prediction, because osmotic stress affects growth rate, which affects the dilution rate of the observed fluorescence.

On the other hand, because good control results are possible using a coarse model of the controlled system, the platform I have presented here should not be limited to the HOG system to induce gene expression. It could be applied to other gene induction systems, provided that a model relating the input to the gene expression output is available. In addition it has to be possible to activate this gene induction system by the microfluidic device, meaning that the system needs to react to pulses on an inducer which cannot be shorter than 5 minutes. In the next chapter I will show how the gene expression control platform can be used with the methionine inducible promoter.

One reason for choosing MPC instead of a simpler control method was that it allows the integration of constraints on the input. This is important for a gene expression control using the HOG cascade, because prolonged activations of the HOG pathway lead to an adaptation response. Here I constrained the osmotic shocks to be separated by at least 20 minutes, which prevents desensitization of the pathway. Note that with a simple proportional-integral (PI) controller it would not have been possible to reach high fluorescence values. This is because a simple PI controller would lead to a situation in which a high stress is maintained over a long time period, which would lead to cell adaptation and eventually to a situation where the maximal stress is applied and the system has adapted to this stress. The reason for this is that a PI controller uses the time-integral of the error to compute the control-input. If the control-input is bounded this can lead to a situation where the control value increases, even though the maximal input is already applied (windup effect). Together with a system that shows adaptation, this leads to the situation mentioned above in which in the maximal input is applied, but the system adapts to this input.

Summing up, I have presented a control platform which can control gene expression in both populations and single cells at high accuracy over a time of 15 hours. The control targets can be both constant or dynamic. I showed that single cell control is more precise compared to the control of a cell within a population and that single cell control reduces gene expression noise. This suggests that the control of single cells might play an important role in understanding the dynamics of cellular systems.

6

Extension of the gene expression control platform

In the previous chapter I presented an integrated platform for the control of gene expression which uses the HOG signaling cascade as a means to activate gene expression. In principle it should be straightforward to adapt the setup I have proposed to use another gene induction system. Ideally this would only necessitate the adaption of the model which predicts the gene expression response and in addition minor changes to the way the control is computed (e.g. different input constraints). Together with the master student Aishah Rumaysa Prastowo we investigated this question in the course of an internship. The work I present in this chapter has been done by Aishah Rumaysa Prastowo under my supervision.

6.1 Development process

We decided to use the methionine repressible promoter of the *MET3* gene (see Section 2.2.1) as a gene induction system. The choice of this system was motivated by the fact that a model simple enough for controlling purposes has already been proposed by Charvin *et al.* [131]. In addition

Gilles Charvin kindly provided us with the yeast strain he had used to generate the data the model is based on. Another option would have been to use a tetracycline controlled promoter, with the advantage that this gene activation system stems from bacteria and is not naturally present in yeast cells, which would have ruled out pleiotropic effects of the MET3 system. These pleiotropic effects amount for example to the increased production of enzymes synthesizing methionine, once methionine is removed from the media, which might affect the overall behavior of the controlled cells.

The strain used in this work carries a Venus fluorescent protein fused to a fast degradation tag under the control of a MET3 promoter (yPH113, see Table 3.1). In [131] the authors used this strain to illustrate the capabilities of their newly developed microfluidic device in long-term experiments. In addition they characterized the MET3 activated gene expression in response to a shift of the cells to media lacking methionine for several durations and used this data to parametrize a simple gene expression model. The model they proposed is a linear two-dimensional switched ODE model comprising only the folded and unfolded fluorescent protein levels (see Figure 37).

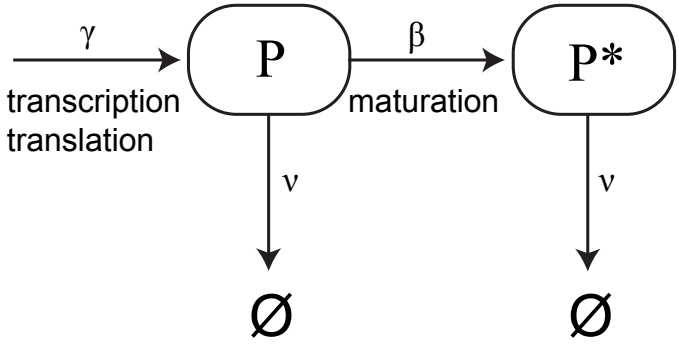


Figure 37: Gene expression model proposed by Charvin *et al.* [131]. The model comprises the unmatured protein level (P) and the matured version (P^*). Unmatured protein is produced with rate γ and then converted in a maturation step to matured protein with rate β. Both, unmatured and matured protein are degraded with a rate proportional to ν.

The equations of this model read

if methionine is present:

$$\dot{P}(t) = -(\beta + \nu)P(t) \tag{6.1}$$

if methionine is not present:

$$\dot{P}(t) = \gamma - (\beta + \nu)P(t) \tag{6.2}$$

$$\dot{P^*}(t) = \beta P(t) - \nu P^*(t) \tag{6.3}$$

with P the level of the unfolded protein and P^* the folded protein. If no methionine is present, P is transcribed with rate γ. The maturation step, in which unfolded protein P folds into P^*, is modeled by a first order reaction with rate β and both P and P^* are degraded with a rate proportional to ν. Charvin *et al.* parametrized the model by fitting it to data produced by transferring cells for defined durations to media lacking methionine in order to activate the MET3 promoter. In addition they used cycloheximide which blocks protein production, to infer the folding rate of the Venus fluorescent protein.

Because we will use this model in control experiments that last 15 hours, the model needs to be able to predict the expression response for longer periods than observed by Charvin *et al.* We investigated the protein production dynamics of this system by alternating the cellular media between media lacking methionine (on) and media containing methionine (off, 200 mg/L). In each experiment, the time with methionine equaled the time without (symmetric input), and the switching times were 15, 30, 45 and 90 minutes respectively. We observed that for all conditions, apart from the 90 minutes input, fluorescence levels oscillate around a fixed value with oscillations induced by the on-durations (see Figure 38). For the 90 minute symmetric input, the overall fluorescence increases over the 15 hours period. Next we tried to identify the model parameters using data from the different experiments with the global optimization procedure CMAES (see Section 2.4). Probably because of the simplicity of the model we were not able to identify a parameter set with which the model is able to reproduce all experiments (see Figure 38). Taking this into account we decided to use model parameters that have been fitted only to the 45 minutes pulse experiment (see Figure 39) and we constrained our control input to pulses of 45 minute duration.

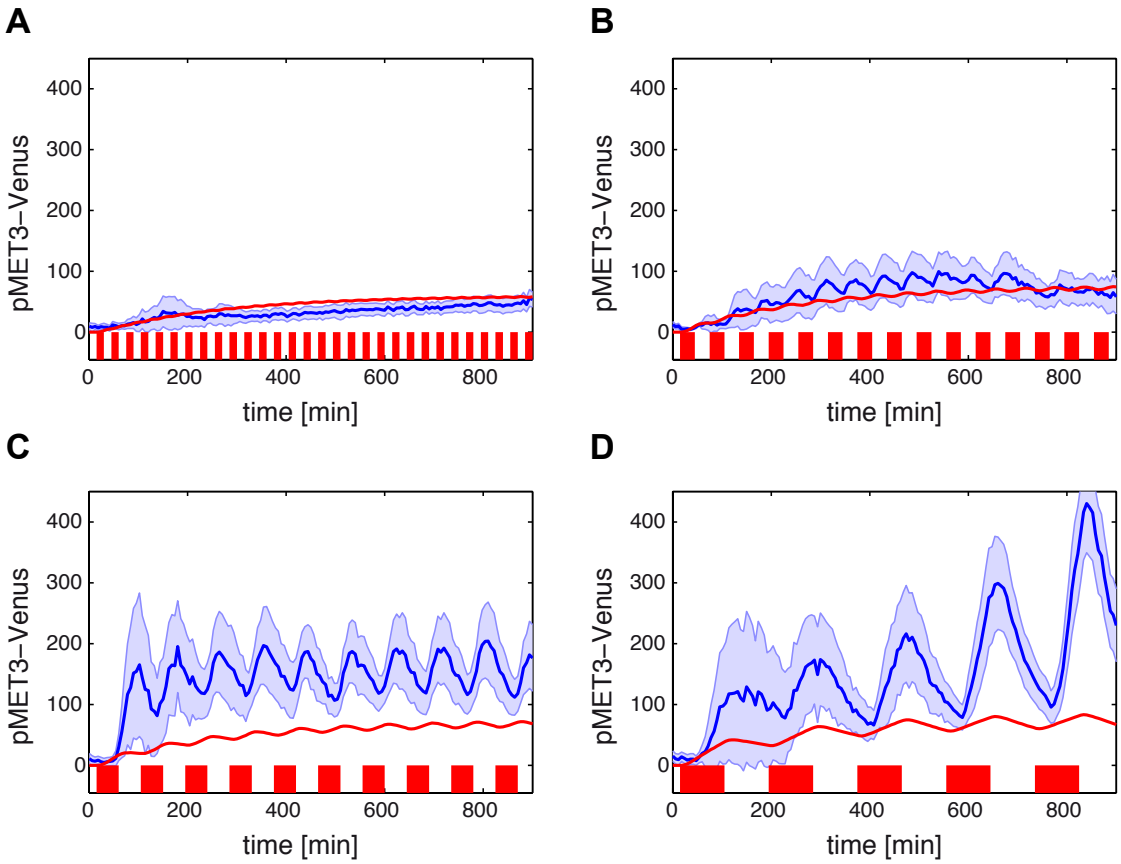


Figure 38: Calibration of pMET3-Venus gene expression and model fit. The cellular media is alternated between media without methionine (on, red bars) and media containing methionine (off). The on and off phases lasted 15 **(A)**, 30 **(B)**, 45 **(C)** and 90 **(D)** minutes. The mean cellular Venus fluorescence is shown as a blue line and the fluorescence standard deviation as a blue envelope. The model shown in Figure 37 was fitted to all four experiments simultaneously and the resulting model simulations are indicated as red lines.

An additional constraint on the input was that after each on-duration (the duration with media lacking methionine), we imposed a minimal period of 25 minutes in which no other on-duration is applied.

The control strategy we applied was very similar to the one described in the previous chapter. The model predictive controller identified the best control strategy by simulating the model for different input profiles for the upcoming 200 minutes. This search was done using CMAES by incorporating a constraint preventing two consecutive on-pulses separated by less than 25 min-

utes. This optimization was only done if a pulse at the current time was possible meaning that the last pulse had ended at least 25 minutes ago. The system state was estimated using an extended Kalman filter (see Section 2.3.3) and the experimental conditions were the same as for the control using the HOG cascade (an observation every 6 minutes). As for the control using the HOG system, the observed time-lag was dealt with by shifting the input by 25 minutes.

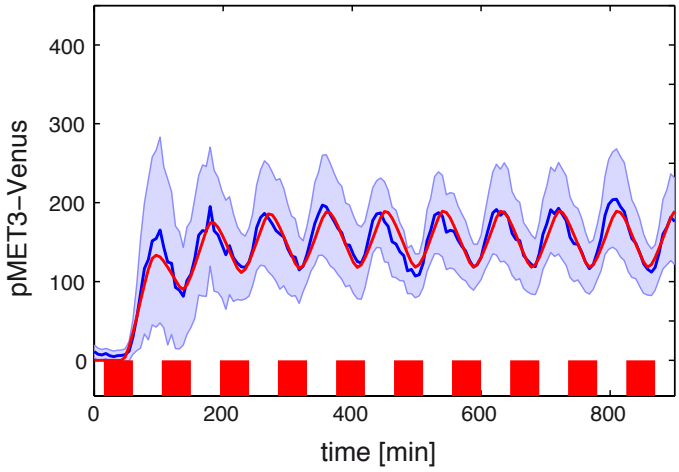


Figure 39: Model fitted only to the gene expression experiment shown in Figure 38C with on and off durations of 45 minutes. Due to its simplicity, the model is not able to reproduce all gene expression data, but a single experiment can be fitted.

6.2 Results

We used this controller to drive protein expression in a population of cells to a fixed target value of 200 f.u. (set-point experiment). The result displayed in Figure 40 shows that the control works in principle, but the long on-period of 45 minutes causes large oscillations around the reference state. In addition the target value 200 f.u. is similar to the mean value reached in the 45 minute calibration experiment (see Figure 39), which suggests that much higher fluorescence values cannot be reached without increasing the on-duration. A possible improvement of this control setup would be to allow for varying on-durations, which would necessitate the development of a model able to predict the gene expression response to various on-durations. This should limit the oscil-

lations by allowing for shorter pulses and also to reach higher expression levels, by using longer pulses. On the other hand we have demonstrated, that it is possible to transpose the gene expression control platform presented in the previous chapter to another gene induction system. This requires the development of a model of the system used to induce gene expression. In this case such a model was already available, but still it was required to calibrate the model using our data.

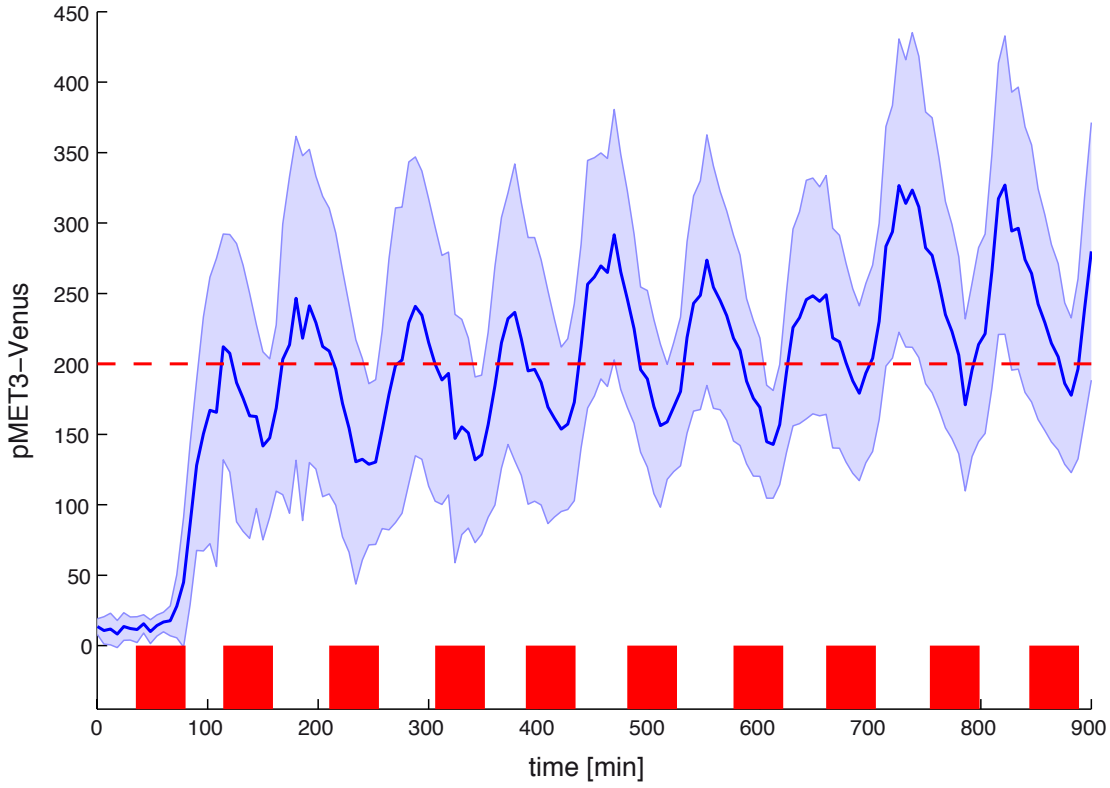


Figure 40: Gene expression control experiment using the MET3 promoter. The target value was set to 200 f.u. and the input was constrained to on-times lasting 45 minutes and a minimal relaxation period of 25 minutes between two consecutive activations. The model predictive controller used the model shown in Figure 37 with the parametrization derived from the calibration experiment shown in Figure 39. The feedback control setup is functional even though the long activation duration of 45 minutes introduces oscillations around the target value.

7 Discussion

7.1 Summary

In this thesis I described the development of a closed-loop control platform for gene expression, which is based on the HOG pathway that is exploited as a system to activate gene expression. My first step towards controlling gene expression, was to develop a control platform to drive the activation of the HOG signaling cascade, following the initial assumption that precise activation of the HOG cascade is a prerequisite for the HOG based control of gene expression. This work showed that real-time control of a cellular process is feasible, even though controlling Hog1 activation to constant target values was not feasible for prolonged times, because of the feedback mechanisms triggered by the HOG pathway. On the other hand I was able to control the activation of Hog1 with a target function consisting of repeated trapeze motifs. Because in this case in between consecutive Hog1 activations there was no osmotic input, the natural feedback response (adaptation) caused normally by activation of the HOG pathway did not take place and repeated activations of the HOG cascade were possible. This result guided the development of

the gene expression control platform. It showed that an amplitude modulation of Hog1 activity is not possible, whereas a frequency encoded Hog1 signal seemed feasible.

Taking this result into account, I developed a gene expression control strategy, which separated the gene expression control problem into two simpler problems, in which in the first step a Hog1 profile leading to the desired gene expression profile was searched for, while in a second step the osmolarity profile implementing the Hog1 profile found in the first round was identified. The Hog1 profiles were constrained to be trapeze motifs with a maximal duration, thereby implementing a frequency encoded Hog1 signal. To prevent the natural feedback response of the HOG system, another constraint ensured a minimal relaxation time between consecutive Hog1 activations. The controller was based on a model predictive control (MPC) approach, incorporating a minimal model of the Hog pathway and of the gene expression response. Computer simulations indicated that this approach would yield a functional gene expression controller for both fixed and varying targets.

Building on this result I started the implementation of a gene expression control platform *in vivo*. Two problems arose during the implementation of the control strategy proposed earlier. First, the microfluidic device I used for the control of the HOG cascade was not amenable for experiments lasting longer than 2-3 hours. This required the use of a different microfluidic device, which allows for long time experiments, but has the drawback of rather slow switching dynamics. This prevented the emulation of analog control by fast switches between high and low sorbitol concentrations, as I had used it for the Hog1 pathway control exploiting the low pass filter behavior of the HOG cascade. For this reason I was not able to apply the proposed input strategy with a continuous osmolarity profile, but instead had to work with an input that is either on (media with 1M sorbitol) or off (normal media). The second problem was that I was not able to monitor the localization of Hog1 over prolonged time periods, because the high imaging frequency required to capture Hog1 dynamics led to bleaching of the fluorescent marker tagged to Hog1. Instead I developed a controller which required the gene expression output as the only measurement. The controller was also based on a MPC approach, but because only one variable was observed (pSTL1-yECitrine fluorescence), I used a simpler model than for the computational investigation. Despite the simplicity of the model, the controller was able to drive gene expression with quantitative accuracy over a time period of at least 15 hours. Both time-constant and time-varying target profiles were achieved and the controller was able to drive gene expression

within a 10-fold range, illustrating the versatility of the approach. In addition to controlling a cell population, the controller was also able to control gene expression in a single cell over 15 hours. I showed that controlling a single cell gives more precision than population control, when the controller reacted directly to the behavior of the controlled cell. Single cell control could for example be important in cases where a large cell to cell heterogeneity limits the precision of population control. Implementing a similar control system, which uses the methionine repressible promoter instead of the HOG pathway to activate gene expression, showed that the platform is extensible to use other gene induction systems without major modifications.

7.2 Limitations of the approach

The main limitation of the control platform in its current state is that it only controls the fluorescence level and not the actual protein level. These two entities differ because each observation bleaches a proportion of the fluorescent protein, which is thereafter still present but cannot be observed. The most obvious option would be to model bleaching and use this to calculate back the total protein amount which is then used for control. Since the way bleaching could be modeled here is not indisputable, I decided to control only fluorescence levels in this first platform. A way to validate a bleaching model would be to fuse two fluorescent proteins with different colors together and then to take images of the second fluorophore sporadically to validate the model. Another option to circumvent the bleaching problem would be to use a gene expression readout that does not suffer from bleaching, for example the light emitting enzyme luciferase. It has to be noted though, that such an approach would lead to other problems, for example the weak signal of luciferases [146]. A third option would be to limit bleaching, either by applying less light in each observation or by using a protein in which the degradation rate dominates the bleaching rate.

A question is whether the control performance could be improved by improving the model. The current model does not capture the full dynamics of the system, which is not surprising, given that a complex signaling cascade and the downstream gene expression process are modeled by a two-dimensional ODE system. Because the state of the model has to be estimated based only on fluorescence observations, it remains unclear whether a more detailed model could improve the control performance. The inability of the current model to capture all observed system dynamics is illustrated by the deviations between open-loop control experiments and the corresponding model simulations (see Figure 32).

Lastly, as seen in Section 2.1.3, activation of the HOG cascade triggers a major transcriptional response with more than 600 induced genes, wherefore using this system to control gene expression in real life applications will have pleiotropic effects. These pleiotropic effects of the Hog1 induced gene expression could be circumvented by using an orthogonal gene induction system which has a less severe impact on cellular processes. One option would be the tet-on system described in Section 2.2.2, which has the advantage that it stems from bacteria and therefore has almost no pleiotropic effects in yeast. Another option would be to use an optogenetic system, for example the phy-GBD/pif-GAD system (see Section 2.2.3 and 7.4).

7.3 Future directions

Despite the limitations mentioned above, the device I presented here offers unprecedented opportunities to decipher the dynamics of gene networks, because it allows for the first time to apply precise time-varying perturbations to the level of a protein. Currently, most perturbations of gene networks are static, but time-varying perturbations provide more information about the perturbed systems [27, 37]. For example, the expression of a transcription factor could be controlled and the expression of the gene downstream of this transcription factor could be observed. By forcing the transcription factor to follow oscillatory expression profiles with different frequencies, the dynamic capabilities of the gene network could be analyzed as it has been done for signaling cascades [38, 39].

Another possible use of the platform would be for optimal experimental design. Apgar *et al.* used a model predictive controller (MPC) to discriminate between different model variants [26]. The idea was to fix a target profile and then use the different models together with a model predictive controller to find inputs signals that when applied to the system lead to the previously defined target profile. The reasoning was that the model which describes the modeled process most accurately, would generate an input profile that when applied to the real system would yield a trajectory similar to the defined target. Apgar *et al.* show the applicability of this approach *in silico*. With the platform I have presented here, this approach could be driven even one step further: instead of computing the input signal before the experiment in an open-loop manner, the model variants could instead be used to drive the closed loop control platform. The model which then would yield the best control performance could be characterized as the model describing the underlying process with the best accuracy. One problem here, which is also present in the open

loop approach presented by Apgar *et al.*, is that this approach might be very sensitive to different parametrizations of the models.

The platform in its current form allows one only to turn on gene expression, while no influence of the degradation rate of the controlled protein is possible. Protein degradation rates can be influenced for example by using the auxin based degradation system from plants [147] in which the plant hormone auxin induces the activity of an ubiquitin ligase, which recognizes a specific amino acid sequence. This would require to express this specific ubiquitin ligase in the cell and to fuse a recognition sequence to the controlled protein. Another method makes use of a cryptic degron[1], which is only exposed in the presence of a small ligand molecule, which then can be used to activate protein degradation of a specific protein [148]. Being able to induce protein degradation would vastly improve the dynamic range of feasible target profiles.

7.4 Related work

First successful applications of control theory to biology were the studies of biological systems using methods from control engineering. David Angeli and Eduardo Sontag identified a class of biological models that allows an elegant analysis of features like oscillatory behavior, bistability or multistability and bifurcations using control theory methodologies. They view a biological system as an input/output (I/O) control system, in which an input u can be applied and one variable is defined as an output. The systems they study have to fulfil two conditions, (i) monotonicity and (ii) monostability. Simply speaking a system is monotone, if all cycles in the incidence graph have a positive sign. The incidence graph represents the direct influence of one species on another and can be constructed knowing only the signs of the elements of the Jacobian matrix. Monostability means that for any constant input the system converges to a stable steady state. If a system is both monotone and monostable, it does not show stable oscillatory behavior and it is possible to determine whether it is bistable in an elegant way that does not rely on numerical simulations [149]. In addition, it can be possible to decompose a non-monotonic system into multiple monotone subsystems, thereby allowing to analyze feedback loops, which are by definition not monotone [150]. This led for example to the development of minimal gain theorems, conditions sufficient to prevent oscillations, in MAP kinase cascades [151].

[1] A degron is a amino acid sequence within a protein that promotes degradation.

Several people studied the robustness of biological systems using control approaches. Yi *et al.* analyzed the adaptation feature of bacterial chemotaxis from a control perspective and concluded that the chemotaxis system achieves robust adaptation through integral feedback control [152]. El-Samad *et al.* studied the heat shock response in *E. coli* and found that the complexity of the system is necessary to achieve features like robustness to parameter variations, noise rejection, speed of response or the economical use of resources [153], features that also engineered systems are optimized for. Recently, Chandra *et al.* analyzed oscillations that occur within the glycolysis and came to the conclusion that these oscillations are not a feature of the system, but rather a product of a trade-off between robustness and efficiency of the glycolysis [154].

Control theory has also been utilized to optimize the application of drugs against HIV [155], tuberculosis [156] or diabetes [157].

Recently, several frameworks have been presented that use feedback regulation to control a biological process, in a way similar to the work presented in this thesis. A framework similar to the one I use has been published by Menolascina *et al.* [158]. They present a computer simulation of a system controlling a synthetic gene network in yeast. Similar to my work, they assume activation of the system by adding an inducer to the cellular media using a microfluidic device and observation using fluorescent proteins. While I use a model predictive control approach, Menolascina *et al.* use a PI controller equipped with a Smith predictor to account for the time delay of the system. In computer simulations they were able to control a model of their system to follow both fixed and dynamic target profiles. Preliminary experimental results showed the feasibility of this approach.

Toettcher *et al.* implemented a feedback control system in live mammalian cells using advanced optogenetic techniques, which is able to control the localization of a signaling protein for both constant and time-varying target profiles [41]. They used the optogenetic phy-PIF module (see Section 2.2.3) which allows one to control the association of the two proteins phy and PIF by applying red or infrared light. Red light activates the binding between the two proteins by inducing a conformational change of phy, whereas infrared light deactivates the binding. By using a digital mirror device they were able to apply light pulses to sub-cellular regions, thereby stimulating each cell individually. They fixed phy to the cellular membrane and labeled PIF with a blue fluorescent protein (BFP), so that they were able to quantify the association state of phy and PIF, by monitoring the membrane localization of PIF-BFP. Using a PI controller, they were able

to control the membrane association of PIF for both constant and time-varying target profiles in single cells. The parameters of the PI controller were tuned using a mathematical model of the phy-PIF system and the authors showed that for high gains the control system shows oscillations. The mathematical model has been tuned by calibration experiments which monitored the steady state response to various constant input signals. With their approach, the authors were able to reduce the cell-to-cell variability of the phy-PIF response, by applying different inputs to each cell.

By fusing phy and PIF to effector proteins of which the activity can be controlled by their localization, this system permits the control of downstream processes. The authors used this to control the activity of the signaling protein phosphoinositide-3 kinase (PI3K), which produces 3′-phosphoinositides (PIP_3) that are involved in controlling cellular functions such as growth, proliferation or differentiation [159]. They fused a PI3K binding domain to PIF, which allowed them to recruit PI3K to the plasma membrane by activation of the phy-PIF module. Since the substrates of PI3K are located within the plasma membrane they were able to control the activity of PI3K by controlling its localization. Using this system they were able to clamp the concentration of the product of PI3K (PIP_3).

The approach of Toettcher *et al.* allows the precise control of the localization of a protein, which permits the activation of signaling proteins that can be stimulated by their localization. Their control systems reacts very fast (within several seconds), which allows the control of fast processes. On the other hand, the approach is less suited for long time experiments, because the high imaging frequency required would induce bleaching of the fluorescent proteins necessary to monitor the response of the system. A problem that is not discussed in their manuscript is how the cellular contours are detected in an image. Because their experiments lasted typically less than a minute such that cellular movements can be neglected, a manual detection which requires human interaction seems likely. Another limitation of their approach is that it requires advanced light application techniques which renders it difficult for other people to use their system. Despite these limitations, the control platform of Toettcher *et al.* is an ideal tool to investigate the dynamics of signaling processes, because it allows the precise and fast activation of signaling cascades that can be activated by the localization of a protein.

The most closely related work has been published by Milias-Argeitis *et al.*, who also used optogenetics to control the expression of a yeast gene to a constant target value over several hours in a

population of cells [42]. They use the phy-GBD/PIF-GAD gene activation system (see Section 2.2.3) in which the association of phy and PIF controls the association of a GAL4 binding domain and a GAL4 activation domain [18]. This allows one to activate the expression of a yellow fluorescent protein (YFP) using red light and inactivate its expression using infrared light. The input was applied using a custom-built light pulse delivery system which can be controlled via a computer. Measurements were taken every 30 minutes using a flow cytometer. This step required manual interaction to deliver an aliquot to the flow cytometer. The authors developed a switched linear ODE model consisting of 4 states and 5 parameters to model their system and fitted the parameters using calibration experiments. Similar to my work, they use a model predictive control strategy, which minimizes the deviation between the model output and a target value and then applies the best control strategy identified in this simulation. The controller was allowed to apply an input pulse (either ON, OFF or do nothing) at discrete times every 15 minutes, while the prediction horizon was 90 minutes. This allowed them simulate all possible input strategies in between a measurement and the next control input, without using an optimization algorithm. A Kalman filter was used to estimate the state of the system and the authors tuned it using *in silico* simulations. This setup allows the authors to control YFP fluorescence to a fixed target value for about 7 hours.

The authors control gene expression of a large population of cells in liquid media. Therefore, even if the volume they use is very small (120 μL) and scaling up laboratory experiments to industry scale has proven difficult [160], their approach seems suited for many biotechnological chemostat applications. The controller Milias-Argeitis *et al.* present works well and shows a high precision. In particular the authors show that they are able to control the system to a fixed setpoint after they have sent a random series of pulses. The model which is used to predict the behavior of the system describe the system well, as the authors show a high agreement between experimental results and model predictions. Nonetheless open-loop control fails in this setup, which again highlights the need of closed-loop control approaches to control gene expression.

In it's present state, a limitation of the approach of Milias-Argeitis *et al.* is that it requires manual interactions for each measurement, which limits the duration of control experiments. This may explain why the control experiments which are presented by the authors show only fixed control targets, even if their approach is in principle also suited for time-varying target profiles. Measuring fluorescence using a flow cytometer has the advantages that noise cancels out because of the

large number of cells that are observed and that full distributions of the fluorescence in the cell population are obtained. However this approach does not allow for single-cell tracking, which is required to analyze and control gene expression in single cells.

7.5 Final remarks

Feedback control of biological processes *in vivo* has just emerged recently as I have detailed in the previous section. On the other hand methods from control theory play an important role in understanding biological systems, because they often resemble a feedback control systems (see Section 1.1.1). Several studies highlighted the importance feedback control to understand biological systems [161–163]. It is likely that *in vivo* feedback control will play a critical role in the application of precise perturbations to biological systems, because the ability of applying precise perturbations is a prerequisite to decipher the functioning of complex biological systems. Obtaining quantitative models with current experimentation techniques is extremely difficult. This is illustrated by the work of Cantone *et al.*, who implemented a synthetic gene network in yeast cells and then constructed a mathematical model for this network [164]. Despite the use of well characterized genes, the production of high quality quantitative data and the use of state of the art modeling and parameter search methods, construction of a quantitative model has been extremely challenging and the authors claim their model to be "semiquantitative", meaning that the model fails to capture all system dynamics in a quantitative manner. Observing the behavior of this network in more diverse situations might help to model it.

Time-varying protein perturbations could also give new insights into cellular processes that depend on oscillatory protein levels. Charvin *et al.* were able to drive cell cycle progression in yeast cells, by controlling the expression of a cyclin [131]. Coudreuse *et al.* recently constructed a minimal cell cycle network in fission yeast, which is driven by the oscillation of a single cyclin-dependent protein kinase [165]. Controlling precisely the frequency and amplitude of proteins involved in the cell cycle will open new ways to investigate the functioning this complex system.

In addition to controlling gene expression, the platform I presented here can be used for real-time experiments, in which the stimulus applied to the cells depends on a readout of the cellular behavior. This would require only slight modifications to the control code I have developed. An example would be on-line parameter learning, where parameter identification and the corresponding biological experiments would be coupled. The benefit here would be that inputs could

be selected, such that they maximize the information gained about parameters that are not well defined in the current parametrization of the model. This approach could also be integrated in the model predictive control approach I have presented, such that the model used for control is constantly updated based on the experimental results.

It has to be noted, that the work I presented here should be seen as a proof-of-concept implementation of a feedback control system for gene expression. But even though significant work remains to actually apply feedback control systems for the generation of precise perturbations of biological systems, this approach might have formidable implications in the future. The recent success of systems biology illustrates the importance of investigating not only the structure and function, but also the dynamics of biological systems. In this context it might very well be the case that feedback control becomes a tool as widely used in quantitative biology as the polymerase chain reaction (PCR) is used nowadays in genomic research. Feedback control systems could be applied to investigate the input-output behavior of biological systems and real-time experimental setups would in addition allow one to do this is a semiautomatic fashion.

Listing of figures

References

[1] A. Böhm and W. Boos. Gene regulation in prokaryotes by subcellular relocalization of transcription factors. *Current opinion in microbiology*, 7(2):151–156, 2004.

[2] J. Monod and F. Jacob. General conclusions: teleonomic mechanisms in cellular metabolism, growth, and differentiation. 26:389–401, 1961.

[3] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular biology of the cell*. Garland Science, 2002.

[4] Y. Wan, K. Qu, Z. Ouyang, M. Kertesz, J. Li, R. Tibshirani, D.L. Makino, R.C. Nutter, E. Segal, and H.Y. Chang. Genome-wide measurement of rna folding energies. *Molecular Cell*, 2012.

[5] Steven L Roderick. The< i> lac</i> operon galactoside acetyltransferase. *Comptes Rendus Biologies*, 328(6):568–575, 2005.

[6] M. Freeman. Feedback control of intercellular signalling in development. *Nature*, 408(6810):313–319, 2000.

[7] A. Becskei and L. Serrano. Engineering stability in gene networks by autoregulation. *Nature*, 405(6786):590–593, 2000.

[8] M. Goulian. Robust control in bacterial regulatory circuits. *Current opinion in microbiology*, 7(2):198–202, 2004.

[9] E.M. Ozbudak, M. Thattai, H.N. Lim, B.I. Shraiman, and A. Van Oudenaarden. Multistability in the lactose utilization network of escherichia coli. *Nature*, 427(6976):737–740, 2004.

[10] D.V. Goeddel, D.G. Kleid, F. Bolivar, H.L. Heyneker, D.G. Yansura, R. Crea, T. Hirose, A. Kraszewski, K. Itakura, and A.D. Riggs. Expression in escherichia coli of chemically synthesized genes for human insulin. *Proceedings of the National Academy of Sciences*, 76(1):106, 1979.

[11] H. Alper, C. Fischer, E. Nevoigt, and G. Stephanopoulos. Tuning genetic control through promoter engineering. *Proceedings of the National Academy of Sciences of the United States of America*, 102(36):12678, 2005.

[12] C.M. Jørgensen, K. Hammer, P.R. Jensen, and J. Martinussen. Expression of the pyrg gene determines the pool sizes of ctp and dctp in lactococcus lactis. *European Journal of Biochemistry*, 271(12):2438–2445, 2004.

[13] T. Ellis, X. Wang, and J.J. Collins. Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nature biotechnology*, 27(5):465–471, 2009.

[14] J. Tornøe, P. Kusk, T.E. Johansen, and P.R. Jensen. Generation of a synthetic mammalian promoter library by modification of sequences spacing transcription factor binding sites. *Gene*, 297(1):21–32, 2002.

[15] J. Boyer, G. Badis, C. Fairhead, E. Talla, F. Hantraye, E. Fabre, G. Fischer, C. Hennequin, R. Koszul, I. Lafontaine, et al. Large-scale exploration of growth inhibition caused by over-expression of genomic fragments in saccharomyces cerevisiae. *Genome biology*, 5(9):R72, 2004.

[16] Y. Fujita, H. Tohda, Y. Giga-Hama, and K. Takegawa. Heat shock-inducible expression vectors for use in schizosaccharomyces pombe. *FEMS yeast research*, 6(6):883–887, 2006.

[17] K.D. Scharf, I. Höhfeld, and L. Nover. Heat stress response and heat stress transcription factors. *Journal of biosciences*, 23(4):313–329, 1998.

[18] S. Shimizu-Sato, E. Huq, J.M. Tepperman, P.H. Quail, et al. A light-switchable gene promoter system. *Nature biotechnology*, 20(10):1041–1044, 2002.

[19] C. Berens and W. Hillen. Gene regulation by tetracyclines. *European Journal of Biochemistry*, 270(15):3109–3121, 2003.

[20] E.M. Ozbudak, M. Thattai, I. Kurtser, A.D. Grossman, and A. van Oudenaarden. Regulation of noise in the expression of a single gene. *Nature genetics*, 31(1):69–73, 2002.

[21] M.B. Elowitz, A.J. Levine, E.D. Siggia, and P.S. Swain. Stochastic gene expression in a single cell. *Science Signalling*, 297(5584):1183, 2002.

[22] T.S. Gardner, C.R. Cantor, and J.J. Collins. Construction of a genetic toggle switch in-escherichia coli. *Nature*, 403:339–342, 2000.

[23] M.B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–338, 2000.

[24] J.A. Stapleton, K. Endo, Y. Fujita, K. Hayashi, M. Takinoue, H. Saito, and T. Inoue. Feedback control of protein expression in mammalian cells by tunable synthetic translational inhibition. *ACS Synthetic Biology*, 2011.

[25] A. Polynikis, G. Cuccato, S. Criscuolo, S.J. Hogan, M.D. Bernardo, and D.D. Bernardo. Design and construction of a versatile synthetic network for bistable gene expression in mammalian systems. *Journal of Computational Biology*, 18(2):195–203, 2011.

[26] J.F. Apgar, J.E. Toettcher, D. Endy, F.M. White, and B. Tidor. Stimulus design for model selection and validation in cell signaling. *PLoS Computational Biology*, 4(2):e30, 2008.

[27] J. Ang, B. Ingalls, and D. McMillen. Probing the input-output behavior of biochemical and genetic systems system identification methods from control theory. *Methods Enzymol*, 487:279–317, 2011.

[28] G.W. Li and X.S. Xie. Central dogma at the single-molecule level in living cells. *Nature*, 475(7356):308–315, 2011.

[29] A. Raj and A. Van Oudenaarden. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, 135(2):216–226, 2008.

[30] A. Eldar and M.B. Elowitz. Functional roles for noise in genetic circuits. *Nature*, 467(7312):167–173, 2010.

[31] N.C. Shaner, P.A. Steinbach, and R.Y. Tsien. A guide to choosing fluorescent proteins. *Nature Methods*, 2(12):905–909, 2005.

[32] S. Hohmann. Osmotic stress signaling and osmoadaptation in yeasts. *Microbiology and Molecular Biology Reviews*, 66(2):300–372, 2002.

[33] P.J. Westfall, D.R. Ballon, and J. Thorner. When the stress of your environment makes you go hog wild. *Science's STKE*, 306(5701):1511, 2004.

[34] F. Posas, J.R. Chambers, J.A. Heyman, J.P. Hoeffler, E. de Nadal, and J. Ariño. The transcriptional response of yeast to saline stress. *Journal of Biological Chemistry*, 275(23):17249, 2000.

[35] M. Rep, M. Krantz, J.M. Thevelein, and S. Hohmann. The transcriptional response of saccharomyces cerevisiae to osmotic shock. *Journal of Biological Chemistry*, 275(12):8290–8300, 2000.

[36] Arthur Edelstein, Nenad Amodaj, Karl Hoover, Ron Vale, and Nico Stuurman. Computer control of microscopes using μmanager. *Current Protocols in Molecular Biology*, 2010.

[37] O. Lipan and W.H. Wong. The use of oscillatory signals in the study of genetic networks. *Proceedings of the National Academy of Sciences of the United States of America*, 102(20):7063–7068, 2005.

[38] P. Hersen, M.N. McClean, L. Mahadevan, and S. Ramanathan. Signal processing by the hog map kinase pathway. *Proceedings of the National Academy of Sciences of the United States of America*, pages 7165–7170, 2008.

[39] J.T. Mettetal, D. Muzzey, C. Gómez-Uribe, and A. van Oudenaarden. The frequency dependence of osmo-adaptation in saccharomyces cerevisiae. *Science*, 319:482, 2008.

[40] J. Uhlendorf, S. Bottani, F. Fages, P. Hersen, and G. Batt. Towards real-time control of gene expression: controlling the hog signaling cascade. *Pacific Symposium On Biocomputing*, pages 338–349, 2011.

[41] J.E. Toettcher, D. Gong, W.A. Lim, and O.D. Weiner. Light-based feedback for controlling intracellular signaling dynamics. *Nature Methods*, 8(10):837–839, 2011.

[42] A. Milias-Argeitis, S. Summers, J. Stewart-Ornstein, I. Zuleta, D. Pincus, H. El-Samad, M. Khammash, and J. Lygeros. In silico feedback for in vivo regulation of a gene expression circuit. *Nature biotechnology*, 2011.

[43] J. Uhlendorf, A. Miermont, T. Delaveau, G. Charvin, F. Fages, S. Bottani, G. Batt, and P. Hersen. Long-term model predictive control of gene expression at the population and single-cell levels. *Proceedings of the National Academy of Sciences*, 109(35):14271–14276, 2012.

[44] J.A. Barnett. A history of research on yeasts 1: work by chemists and biologists 1789–1850. *Yeast*, 14(16):1439–1451, 1998.

[45] J.A. Barnett. A history of research on yeasts 8: taxonomy. *Yeast*, 21(14):1141–1193, 2004.

[46] J.A. Barnett. Beginnings of microbiology and biochemistry: the contribution of yeast research. *Microbiology*, 149(3):557–567, 2003.

[47] A. Goffeau, BG Barrell, H. Bussey, RW Davis, B. Dujon, H. Feldmann, F. Galibert, JD Hoheisel, C. Jacq, M. Johnston, et al. Life with 6000 genes. *Science*, 274(5287):546–567, 1996.

[48] E.A. Winzeler, D.D. Shoemaker, A. Astromoff, H. Liang, K. Anderson, B. Andre, R. Bangham, R. Benito, J.D. Boeke, H. Bussey, et al. Functional characterization of the s. cerevisiae genome by gene deletion and parallel analysis. *Science*, 285(5429):901–906, 1999.

[49] G. Giaever, A.M. Chu, L. Ni, C. Connelly, L. Riles, S. Véronneau, S. Dow, A. Lucau-Danila, K. Anderson, B. André, et al. Functional profiling of the saccharomyces cerevisiae genome. *Nature*, 418(6896):387–391, 2002.

[50] S. Ghaemmaghami, W.K. Huh, K. Bower, R.W. Howson, A. Belle, N. Dephoure, E.K. O'Shea, J.S. Weissman, et al. Global analysis of protein expression in yeast. *Nature*, 425(6959):737–741, 2003.

[51] W.K. Huh, J.V. Falvo, L.C. Gerke, A.S. Carroll, R.W. Howson, J.S. Weissman, E.K. O'Shea, et al. Global analysis of protein localization in budding yeast. *Nature*, 425(6959):686–691, 2003.

[52] D. Botstein and G.R. Fink. Yeast: an experimental organism for 21st century biology. *Genetics*, 189(3):695–704, 2011.

[53] T.L. Orr-Weaver, J.W. Szostak, and R.J. Rothstein. Yeast transformation: a model system for the study of recombination. *Proceedings of the National Academy of Sciences*, 78(10):6354, 1981.

[54] H.C. Causton, B. Ren, S.S. Koh, C.T. Harbison, E. Kanin, E.G. Jennings, T.I. Lee, H.L. True, E.S. Lander, and R.A. Young. Remodeling of yeast genome expression in response to environmental changes. *Molecular biology of the cell*, 12(2):323–337, 2001.

[55] M. Kresnowati, WA Van Winden, MJH Almering, A. Ten Pierick, C. Ras, TA Knijnenburg, P. Daran-Lapujade, JT Pronk, JJ Heijnen, and JM Daran. When transcriptome meets metabolome: fast cellular responses of yeast to sudden relief of glucose limitation. *Molecular systems biology*, 2(1), 2006.

[56] J.M. Buescher, W. Liebermeister, M. Jules, M. Uhr, J. Muntel, E. Botella, B. Hessling, R.J. Kleijn, L. Le Chat, F. Lecointe, et al. Global network reorganization during dynamic adaptations of bacillus subtilis metabolism. *Science*, 335(6072):1099–1103, 2012.

[57] C.A. Sellick, R.N. Campbell, and R.J. Reece. Galactose metabolism in yeast–structure and regulation of the leloir pathway enzymes and the genes encoding them. *International review of cell and molecular biology*, 269:111–150, 2008.

[58] A. Miermont, F. Waharte, H. Hu, M. McClean, S. Bottani, S. Léon, and P. Hersen. Severe osmotic compression triggers a global slow-down of intracellular signaling due to increased molecular crowding. *submitted*, 2012.

[59] J.L. Brewster, T. de Valoir, N.D. Dwyer, E. Winter, and M.C. Gustin. An osmosensing signal transduction pathway in yeast. *Science*, 259(5102):1760–1763, 1993.

[60] T. Maeda, SM Wurgler-Murphy, H. Saito, et al. A two-component system that regulates an osmosensing map kinase cascade in yeast. *Nature*, 369(6477):242, 1994.

[61] J. Yale and H.J. Bohnert. Transcript expression in saccharomyces cerevisiae at high salinity. *Journal of Biological Chemistry*, 276(19):15996–16007, 2001.

[62] A.P. Gasch, P.T. Spellman, C.M. Kao, O. Carmel-Harel, M.B. Eisen, G. Storz, D. Botstein, and P.O. Brown. Genomic expression programs in the response of yeast cells to environmental changes. *Molecular Biology of the Cell*, 11(12):4241–4257, 2000.

[63] S. Hohmann. Control of high osmolarity signalling in the yeast saccharomyces cerevisiae. *FEBS letters*, 583(24):4025–4029, 2009.

[64] H. Dihazi, R. Kessler, and K. Eschrich. High osmolarity glycerol (hog) pathway-induced phosphorylation and activation of 6-phosphofructo-2-kinase are essential for glycerol accumulation and yeast cell proliferation under hyperosmotic stress. *Journal of Biological Chemistry*, 279(23):23961–23968, 2004.

[65] M.J. Tamás, K. Luyten, F.C.W. Sutherland, A. Hernandez, J. Albertyn, H. Valadi, H. Li, B.A. Prior, S.G. Kilian, J. Ramos, et al. Fps1p controls the accumulation and release of the compatible solute glycerol in yeast osmoregulation. *Molecular microbiology*, 31(4):1087–1104, 1999.

[66] A. Miermont, J. Uhlendorf, M. McClean, P. Hersen, et al. The dynamical systems properties of the hog signaling cascade. *Journal of signal transduction*, 2011:930940, 2011.

[67] T. Mizuno and S. Mizushima. Signal transduction and gene regulation through the phosphorylation of two regulatory components: the molecular basis for the osmotic regulation of the porin genes. *Molecular microbiology*, 4(7):1077–1082, 2006.

[68] N. Hao, M. Behar, S.C. Parnell, M.P. Torres, C.H. Borchers, T.C. Elston, and H.G. Dohlman. A systems-biology analysis of feedback inhibition in the sho1 osmotic-stress-response pathway. *Current biology*, 17(8):659–667, 2007.

[69] K. Furukawa, Y. Hoshi, T. Maeda, T. Nakajima, and K. Abe. Aspergillus nidulans hog pathway is activated only by two-component signalling pathway in response to osmotic stress. *Molecular microbiology*, 56(5):1246–1261, 2005.

[70] D. Muzzey, C.A. Gómez-Uribe, J.T. Mettetal, and A. Van Oudenaarden. A systems-level analysis of perfect adaptation in yeast osmoregulation. *Cell*, 138(1):160–171, 2009.

[71] E. Klipp, B. Nordlander, R. Krüger, P. Gennemark, and S. Hohmann. Integrative model of the response of yeast to osmotic shock. *Nature biotechnology*, 23(8):975–982, 2005.

[72] D.B. Berry and A.P. Gasch. Stress-activated genomic expression changes serve a preparative role for impending stress in yeast. *Molecular biology of the cell*, 19(11):4580–4587, 2008.

[73] A.P. Capaldi, T. Kaplan, Y. Liu, N. Habib, A. Regev, N. Friedman, and E.K. O'Shea. Structure and function of a transcriptional network activated by the mapk hog1. *Nature genetics*, 40(11):1300–1306, 2008.

[74] P. Gennemark, B. Nordlander, S. Hohmann, and D. Wedelin. A simple mathematical model of adaptation to high osmolarity in yeast. *In silico biology*, 6(3):193–214, 2006.

[75] Z. Zi, W. Liebermeister, and E. Klipp. A quantitative study of the hog1 mapk response to fluctuating osmotic stress in saccharomyces cerevisiae. *PloS one*, 5(3):e9522, 2010.

[76] C. Kühn, KVS Prasad, E. Klipp, and P. Gennemark. Formal representation of the high osmolarity glycerol pathway in yeast. *Genome informatics*, 20:22–83, 2010.

[77] J. Macia, S. Regot, T. Peeters, N. Conde, R. Sole, and F. Posas. Dynamic signaling in the hog1 mapk pathway relies on high basal signal transduction. *Science's STKE*, 2(63):ra13, 2009.

[78] D. Maya, M.J. Quintero, M. de la Cruz Muñoz-Centeno, and S. Chávez. Systems for applied gene control in saccharomyces cerevisiae. *Biotechnology letters*, 30(6):979–987, 2008.

[79] X. Mao, Y. Hu, C. Liang, and C. Lu. Met3 promoter: a tightly regulated promoter and its application in construction of conditional lethal strain. *Current microbiology*, 45(1):37–40, 2002.

[80] K.T. Tamai, X. Liu, P. Silar, T. Sosinowski, and D.J. Thiele. Heat shock transcription factor activates yeast metallothionein gene expression in response to heat and glucose starvation via distinct signalling pathways. *Molecular and cellular biology*, 14(12):8155–8165, 1994.

[81] W. Hillen and C. Berens. Mechanisms underlying expression of tn10 encoded tetracycline resistance. *Annual Reviews in Microbiology*, 48(1):345–369, 1994.

[82] W. Saenger, P. Orth, C. Kisker, W. Hillen, and W. Hinrichs. The tetracycline repressor—a paradigm for a biological switch. *Angewandte Chemie International Edition*, 39(12):2042–2052, 2000.

[83] A. Welman, J. Barraclough, and C. Dive. Tetracycline regulated systems in functional oncogenomics. *Translational Oncogenomics*, 2:17–33, 2007.

[84] B. Eckert and C.F. Beck. Topology of the transposon tn10-encoded tetracycline resistance protein within the inner membrane of escherichia coli. *Journal of Biological Chemistry*, 264(20):11663–11670, 1989.

[85] TN Nguyen, Q.G. Phan, L.P. Duong, K.P. Bertrand, and R.E. Lenski. Effects of carriage and expression of the tn10 tetracycline-resistance operon on the fitness of escherichia coli k12. *Molecular biology and evolution*, 6(3):213–225, 1989.

[86] C. Gatz and P.H. Quail. Tn10-encoded tet repressor can regulate an operator-containing plant promoter. *Proceedings of the National Academy of Sciences*, 85(5):1394, 1988.

[87] F. Yao, T. Svensjö, T. Winkler, M. Lu, C. Eriksson, and E. Eriksson. Tetracycline repressor, tetr, rather than the tetr-mammalian cell transcription factor fusion derivatives, regulates inducible gene expression in mammalian cells. *Human gene therapy*, 9(13):1939–1950, 1998.

[88] J. Ohkawa and K. Taira. Control of the functional activity of an antisense rna by a tetracycline-responsive derivative of the human u6 snrna promoter. *Human gene therapy*, 11(4):577–585, 2000.

[89] M. Gossen and H. Bujard. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proceedings of the National Academy of Sciences*, 89(12):5547, 1992.

[90] E. Garí, L. Piedrafita, M. Aldea, and E. Herrero. A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in saccharomyces cerevisiae. *Yeast*, 13(9):837–848, 1998.

[91] M. Gossen, S. Freundlieb, G. Bender, G. Muller, W. Hillen, and H. Bujard. Transcriptional activation by tetracyclines in mammalian cells. *Science*, 268(5218):1766–1769, 1995.

[92] S. Urlinger, U. Baron, M. Thellmann, M.T. Hasan, H. Bujard, and W. Hillen. Exploring the sequence space for tetracycline-dependent transcriptional activators: novel mutations yield expanded range and sensitivity. *Proceedings of the National Academy of Sciences*, 97(14):7963, 2000.

[93] G. Bellí, E. Garí, L. Piedrafita, M. Aldea, and E. Herrero. An activator/repressor dual system allows tight tetracycline-regulated gene expression in budding yeast. *Nucleic acids research*, 26(4):942–947, 1998.

[94] D. Nevozhay, R.M. Adams, K.F. Murphy, K. Josić, and G. Balázsi. Negative autoregulation linearizes the dose–response and suppresses the heterogeneity of gene expression. *Proceedings of the National Academy of Sciences*, 106(13):5123–5128, 2009.

[95] J.J. Tabor, A. Levskaya, and C.A. Voigt. Multichromatic control of gene expression in escherichia coli. *Journal of molecular biology*, 405(2):315–324, 2011.

[96] B.V. Zemelman, G.A. Lee, M. Ng, and G. Miesenböck. Selective photostimulation of genetically charged neurons. *Neuron*, 33(1):15–22, 2002.

[97] E.S. Boyden, F. Zhang, E. Bamberg, G. Nagel, and K. Deisseroth. Millisecond-timescale, genetically targeted optical control of neural activity. *Nature neuroscience*, 8(9):1263–1268, 2005.

[98] R.D. Airan, K.R. Thompson, L.E. Fenno, H. Bernstein, and K. Deisseroth. Temporally precise in vivo control of intracellular signalling. *Nature*, 458(7241):1025–1029, 2009.

[99] L. Fenno, O. Yizhar, and K. Deisseroth. The development and application of optogenetics. *Annual review of neuroscience*, 34:389–412, 2011.

[100] P. Hegemann and A. Möglich. Channelrhodopsin engineering and exploration of new optogenetic tools. *nature methods*, 8(1):39–42, 2010.

[101] F. Zhang, L.P. Wang, M. Brauner, J.F. Liewald, K. Kay, N. Watzke, P.G. Wood, E. Bamberg, G. Nagel, A. Gottschalk, et al. Multimodal fast optical interrogation of neural circuitry. *Nature*, 446(7136):633–639, 2007.

[102] S.Q. Lima and G. Miesenböck. Remote control of behavior through genetically targeted photostimulation of neurons. *Cell*, 121(1):141–152, 2005.

[103] A.R. Adamantidis, F. Zhang, A.M. Aravanis, K. Deisseroth, and L. De Lecea. Neural substrates of awakening probed with optogenetic control of hypocretin neurons. *Nature*, 450(7168):420–424, 2007.

[104] X. Han, X. Qian, J.G. Bernstein, H. Zhou, G.T. Franzesi, P. Stern, R.T. Bronson, A.M. Graybiel, R. Desimone, and E.S. Boyden. Millisecond-timescale optical control of neural dynamics in the nonhuman primate brain. *Neuron*, 62(2):191, 2009.

[105] R. Khanna, E. Huq, E.A. Kikis, B. Al-Sady, C. Lanzatella, and P.H. Quail. A novel molecular recognition motif necessary for targeting photoactivated phytochrome signaling to specific basic helix-loop-helix transcription factors. *The Plant Cell Online*, 16(11):3033–3044, 2004.

[106] J.E. Toettcher, C.A. Voigt, O.D. Weiner, and W.A. Lim. The promise of optogenetics in cell biology: interrogating molecular circuits in space and time. *Nature methods*, 8(1):35–38, 2010.

[107] L. Hood, J.R. Heath, M.E. Phelps, and B. Lin. Systems biology and new technologies enable predictive and preventative medicine. *Science Signalling*, 306(5696):640, 2004.

[108] H. Kitano. Systems biology: a brief overview. *Science*, 295(5560):1662–1664, 2002.

[109] M.J. Herrgård, N. Swainston, P. Dobson, W.B. Dunn, K.Y. Arga, M. Arvas, N. Blüthgen, S. Borger, R. Costenoble, M. Heinemann, et al. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nature biotechnology*, 26(10):1155–1160, 2008.

[110] M. Kanehisa, S. Goto, Y. Sato, M. Furumichi, and M. Tanabe. Kegg for integration and interpretation of large-scale molecular data sets. *Nucleic acids research*, 40(D1):D109–D114, 2012.

[111] J.R. Karr, J.C. Sanghvi, D.N. Macklin, M.V. Gutschow, J.M. Jacobs, B. Bolival, N. Assad-Garcia, J.I. Glass, and M.W. Covert. A whole-cell computational model predicts phenotype from genotype. *Cell*, 150(2):389–401, 2012.

[112] E.D. Sontag. *Mathematical control theory: deterministic finite dimensional systems*, volume 6. Springer, 1998.

[113] A.H. Jazwinski. *Stochastic processes and filtering theory*, volume 63. Academic press, 1970.

[114] N. Hansen and A. Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary computation*, 9(2):159–195, 2001.

[115] Domitille Heitzler, Guillaume Durand, Nathalie Gallay, Aurélien Rizk, Seungkirl Ahn, Jihee Kim, Jonathan D Violin, Laurence Dupuy, Christophe Gauthier, Vincent Piketty, et al. Competing g protein-coupled receptor kinases balance g protein and β-arrestin signaling. *Molecular Systems Biology*, 8(1), 2012.

[116] Florine Dupeux, Julia Santiago, Katja Betz, Jamie Twycross, Sang-Youl Park, Lesia Rodriguez, Miguel Gonzalez-Guzman, Malene Ringkjøbing Jensen, Natalio Krasnogor, Martin Blackledge, et al. A thermodynamic switch modulates abscisic acid receptor sensitivity. *The EMBO journal*, 2011.

[117] Klaus Maier, Ute Hofmann, Matthias Reuss, and Klaus Mauch. Dynamics and control of the central carbon metabolism in hepatoma cells. *BMC systems biology*, 4(1):54, 2010.

[118] C. Wählby. *Algorithms for applied digital image cytometry*. PhD thesis, Uppsala University, 2003.

[119] R.O. Duda and P.E. Hart. Use of the hough transformation to detect lines and curves in pictures. *Communications of the ACM*, 15(1):11–15, 1972.

[120] D.H. Ballard. Generalizing the hough transform to detect arbitrary shapes. *Pattern recognition*, 13(2):111–122, 1981.

[121] M. Kvarnström, K. Logg, A. Diez, K. Bodvard, and M. Käll. Image analysis algorithms for cell contour recognition in budding yeast. *Optics express*, 16(17):12943–12957, 2008.

[122] D.W. Paglieroni. Distance transforms: Properties and machine vision applications. *CVGIP: Graphical Models and Image Processing*, 54(1):56–74, 1992.

[123] G. Dantzig. *Linear programming and extensions*. Princeton university press, 1998.

[124] A.Y. Fu, C. Spence, A. Scherer, F.H. Arnold, S.R. Quake, et al. A microfabricated fluorescence-activated cell sorter. *Nature biotechnology*, 17(11):1109–1111, 1999.

[125] C. Zhang, J. Xu, W. Ma, W. Zheng, et al. Pcr microfluidic devices for dna amplification. *Biotechnology advances*, 24(3):243–284, 2006.

[126] C.C. Lee, T.M. Snyder, and S.R. Quake. A microfluidic oligonucleotide synthesizer. *Nucleic acids research*, 38(8):2514–2521, 2010.

[127] S. Paliwal, P.A. Iglesias, K. Campbell, Z. Hilioti, A. Groisman, and A. Levchenko. Mapk-mediated bimodal gene expression and adaptive gradient sensing in yeast. *Nature*, 446(7131):46–51, 2007.

[128] M.R. Bennett, W.L. Pang, N.A. Ostroff, B.L. Baumgartner, S. Nayak, L.S. Tsimring, and J. Hasty. Metabolic gene regulation in a dynamically changing environment. *Nature*, 454(7208):1119–1122, 2008.

[129] N.R. Glass, R. Tjeung, P. Chan, L.Y. Yeo, and J.R. Friend. Organosilane deposition for microfluidic applications. *Biomicrofluidics*, 5(3):036501, 2011.

[130] C.B. Brachmann, A. Davies, G.J. Cost, E. Caputo, J. Li, P. Hieter, JD Boeke, et al. Designer deletion strains derived from saccharomyces cerevisiae s288c: a useful set of strains and plasmids for pcr-mediated gene disruption and other applications. *Yeast*, 14:115–132, 1998.

[131] G. Charvin, F.R. Cross, and E.D. Siggia. A microfluidic device for temporally controlled gene expression and long-term fluorescent imaging in unperturbed dividing yeast cells. *PLoS One*, 3(1):e1468, 2008.

[132] TE Turner, S. Schnell, K. Burrage, et al. Stochastic approaches for modelling in vivo reactions. *Computational Biology and Chemistry*, 28(3):165–178, 2004.

[133] D.T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of computational physics*, 22(4):403–434, 1976.

[134] E. Klipp, W. Liebermeister, C. Wierling, A. Kowald, H. Lehrach, and R. Herwig. *Systems biology*. Wiley-VCH, 2011.

[135] J. Uhlendorf, P. Hersen, and G. Batt. Towards real-time control of gene expression: in silico analysis. In *The 18th World Congress of the International Federation of Automatic Control (IFAC)*, 2011.

[136] C. Ferreira, F. Van Voorst, A. Martins, L. Neves, R. Oliveira, M.C. Kielland-Brandt, C. Lucas, and A. Brandt. A member of the sugar transporter family, stl1p is the glycerol/h+ symporter in saccharomyces cerevisiae. *Molecular biology of the cell*, 16(4):2068–2076, 2005.

[137] A. Henderson, J. Erales, M.A. Hoyt, P. Coffino, A. Henderson, J. Erales, M.A. Hoyt, and P. Coffino. Dependence of proteasome processing rate on substrate unfolding. *Journal of Biological Chemistry*, 286(20):17495–17502, 2011.

[138] L. Cai, N. Friedman, and X.S. Xie. Stochastic protein expression in individual cells at the single molecule level. *Nature*, 440(7082):358–362, 2006.

[139] F. Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945.

[140] M.A. Fligner and G.E. Policello II. Robust rank procedures for the behrens-fisher problem. *Journal of the American Statistical Association*, 76(373):162–168, 1981.

[141] J Clotet, F Posas, et al. Control of cell cycle in response to osmostress: lessons from yeast. *Methods in enzymology*, 428:63, 2007.

[142] Jignesh H Parmar, Sharad Bhartiya, and KV Venkatesh. Quantification of metabolism in saccharomyces cerevisiae under hyperosmotic conditions using elementary mode analysis. *Journal of industrial microbiology & biotechnology*, pages 1–15, 2012.

[143] Gilad Yaakov, Alba Duch, María García-Rubio, Josep Clotet, Javier Jimenez, Andrés Aguilera, and Francesc Posas. The stress-activated protein kinase hog1 mediates s phase delay in response to osmostress. *Molecular biology of the cell*, 20(15):3572–3582, 2009.

[144] Xavier Escoté, Meritxell Zapater, Josep Clotet, and Francesc Posas. Hog1 mediates cell-cycle arrest in g1 phase by the dual targeting of sic1. *Nature cell biology*, 6(10):997–1002, 2004.

[145] Josep Clotet, Xavier Escoté, Miquel Àngel Adrover, Gilad Yaakov, Eloi Garí, Martí Aldea, Eulàlia De Nadal, and Francesc Posas. Phosphorylation of hsl1 by hog1 leads to a g2 arrest essential for cell survival at high osmolarity. *The EMBO journal*, 25(11):2338–2346, 2006.

[146] G. Choy, S. O Connor, F.E. Diehn, N. Costouros, H.R. Alexander, P. Choyke, S.K. Libutti, et al. Comparison of noninvasive fluorescent and bioluminescent small animal optical imaging. *Biotechniques*, 35(5):1022–1031, 2003.

[147] K. Nishimura, T. Fukagawa, H. Takisawa, T. Kakimoto, and M. Kanemaki. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nature methods*, 6(12):917–922, 2009.

[148] K.M. Bonger, L. Chen, C.W. Liu, and T.J. Wandless. Small-molecule displacement of a cryptic degron causes conditional protein degradation. *Nature chemical biology*, 7(8):531–537, 2011.

[149] D. Angeli, J.E. Ferrell, and E.D. Sontag. Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. *Proceedings of the National Academy of Sciences of the United States of America*, 101(7):1822–1827, 2004.

[150] D. Angeli and E. Sontag. Interconnections of monotone systems with steady-state characteristics. *Optimal control, stabilization and nonsmooth analysis*, pages 135–154, 2004.

[151] E.D. Sontag. Asymptotic amplitudes and cauchy gains: a small-gain principle and an application to inhibitory biological feedback. *Systems & control letters*, 47(2):167–179, 2002.

[152] T.M. Yi, Y. Huang, M.I. Simon, and J. Doyle. Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proceedings of the National Academy of Sciences*, 97(9):4649–4653, 2000.

[153] H. El-Samad, H. Kurata, JC Doyle, CA Gross, and M. Khammash. Surviving heat shock: control strategies for robustness and performance. *Proceedings of the National Academy of Sciences of the United States of America*, 102(8):2736–2741, 2005.

[154] F.A. Chandra, G. Buzi, and J.C. Doyle. Glycolytic oscillations and limits on robust efficiency. *science*, 333(6039):187–192, 2011.

[155] JAM Felippe de Souza, M.A.L. Caetano, and T. Yoneyama. Optimal control theory applied to the anti-viral treatment of aids. In *Decision and Control, 2000. Proceedings of the 39th IEEE Conference on*, volume 5, pages 4839–4844. IEEE, 2000.

[156] E. Jung, S. Lenhart, and Z. Feng. Optimal control of treatments in a two-strain tuberculosis model. *Discrete and Continuous Dynamical Systems Series B*, 2(4):473–482, 2002.

[157] R.S. Parker, F.J. Doyle III, and N.A. Peppas. The intravenous route to blood glucose control. *Engineering in Medicine and Biology Magazine, IEEE*, 20(1):65–73, 2001.

[158] F. Menolascina, M. Di Bernardo, and D. Di Bernardo. Analysis, design and implementation of a novel scheme for in-vivo control of synthetic gene regulatory networks. *Automatica*, 47(6):1265–1270, 2011.

[159] B. Vanhaesebroeck, S.J. Leevers, K. Ahmadi, J. Timms, R. Katso, P.C. Driscoll, R. Woscholski, P.J. Parker, and M.D. Waterfield. Synthesis and function of 3-phosphorylated inositol lipids. *Annual review of biochemistry*, 70(1):535–602, 2001.

[160] RS Islam, D. Tisi, MS Levy, and GJ Lye. Scale-up of escherichia coli growth and recombinant protein expression conditions from microwell to laboratory and pilot scale based on matched kla. *Biotechnology and bioengineering*, 99(5):1128–1139, 2008.

[161] M.E. Csete and J.C. Doyle. Reverse engineering of biological complexity. *Science Signalling*, 295(5560):1664, 2002.

[162] H. Kurata, H. El-Samad, R. Iwasaki, H. Ohtake, J.C. Doyle, I. Grigorova, C.A. Gross, and M. Khammash. Module-based analysis of robustness tradeoffs in the heat shock response system. *PLoS computational biology*, 2(7):e59, 2006.

[163] Y.J. Shin and L. Bleris. Linear control theory for gene network modeling. *PloS one*, 5(9):e12785, 2010.

[164] I. Cantone, L. Marucci, F. Iorio, M.A. Ricci, V. Belcastro, M. Bansal, S. Santini, M. di Bernardo, D. di Bernardo, M.P. Cosma, et al. A yeast synthetic network for in vivo assessment of reverse-engineering and modeling approaches. *Cell*, 137(1):172, 2009.

[165] D. Coudreuse and P. Nurse. Driving the cell cycle with a minimal cdk control network. *Nature*, 468(7327):1074–1079, 2010.

A

Description of the control software

In the following, I will give a brief overview over the control software I developed in the course of the gene expression control project. The software is publicly available online[1] and could be re-used for other closed loop control purposes. The automated cell-tracker might be of use for many time-lapse microscopy applications, wherefore I will describe it separately. All software is written in MATLAB.

In addition I have developed a micro-manager driver for the EXFO X-Cite 120 PC fluorescent lamp, because at the time when I started this project no such driver was available. The driver has been integrated into the main micro-manager distribution[2].

A.1 Control software

Here, I will describe the functions and classes that make up the gene expression control software.

`class Model`
The model class contains the ODE description of the controlled system. It provides methods to integrate the model over time.
Functions:

[1] http://www.msc.univ-paris-diderot.fr/~jannis/control_code
[2] htttp://www.micro-manager.org

- `integrate( x_0, on_times, on_durations, time_points )`
  Returns: `[time_interval,x]`
  Integrates the model over time. Inputs are the initial conditions (`x_0`), a vector of times at which the valve is switched to the on-state (`on_times`), a vector of on-durations (`on_durations`) and a vector specifying the time points for numerical simulation (`time_points`). The function returns the time interval vector as well as the solution at each time point (`x`).

`class StateEstimator`
This class implements an extended Kalman filter (see Section 2.3.3).
Functions:

- `StateEstimator( model, x_0, H )`
  Constructor for the class. Inputs are an instance of the `Model` class, the initial conditions (`x_0`) and the observation matrix `H`.

- `estimateState( observation, on_times, on_durations, current_time, time_step)`
  Returns: `x_current`
  Estimates the current state of the model based on observations. Inputs are the observations (`observation`), the past osmotic input (`on_times` and `on_durations`), the current time and the time elapsed since the last estimation (`time_step`).

`class Micromanager`
Class implementing the interface to the microscope via the open source software MicroManager.
Functions:

- `getImage( channel )`
  Returns: `img`
  Function to take an image (at the moment imaging conditions are hard coded). The channel can be specified (e.g. "trans" or "yfp"). The function returns the image as a matrix of gray values (`img`).

`class Valve`
Class implementing the interface to the microfluidic valve.
Functions:

- `switchOn()`
  Switches the valve to the on-state.

- `switchOff()`
  Switches the valve to the off-state.

- `timedSwitch( time_start, on_times, on_durations )` Function to switch the valve in the future. The inputs are the time of the start of the experiment `time_start`, the vector of times at which the valve is to be switched on (`on_times`) and a vector specifying the on-durations (`on_durations`).

`function getOsm( on_times, on_durations, time_interval )`
Function giving the osmolarity profile as it is felt by the cells. This profile is different from the simple state of the valve because of the dead volume in the tubes and because of mixing in the tubes. The profile is computed by a linear interpolation of a calibration measurement (see Figure 20B).

`function find_osm_profile( x_0, time_start, reference, model, on_times, on_durations )`
Returns: `[new_on_times, new_on_durations]`
Function which finds the best future osmolarities to apply in order to make the system follow the target profile. Inputs are the current system state (`x_0`), the current time (`time_start`), the reference profile (`reference`), an instance of the model and the past osmotic input (`on_times` and `on_durations`).

`function evaluate_osm_profile( osmotic_profile, x_0, time_start, time_start_model_sim, reference, model, last_shock_end )`
Returns: `fitness`
Function to evaluate a possible osmotic profile. Inputs are the osmotic profile, the current state of the model (`x_0`), the current time (`time_start`), the time at which the model simulation is to be started (`time_start_model_sim`), the target profile (`reference`), the model instance (`model`) and the time of the end of the last shock (`last_shock_end`). The function returns a fitness value.

The integration of these pieces is done in a script called `do_mpc.m`.

A.2 Cell tracker

The cell-segmentation, tracking and quantification of fluorescent markers is done in the class `HoughTracker`. This class may be used independently of the rest of the control software via the script `do_tracking.m`.

`class HoughTracker`
Functions:

- `HoughTracker( name )`
  Constructor of the class. Output images are saved in the folder `name_tracker`.

- `addTimepoint( img_trans, img_expression, img_hog, img_nucleus )`
  Returns: `sorted_fluorescence`
  Function that does the segmentation and tracking. Inputs are the trans-image (`img_trans`), the fluorescent image of the gene expression marker (`img_expression`), the image of the fluorescence label of Hog1 (`img_hog`) and a fluorescent image of a nuclear marker (`img_nucleus`). All inputs apart from the trans-image are optional and can be set to zero. The function does the tracking and returns a sorted array of the expression values for each detected cell. If a cell has been lost the corresponding index in `sorted_fluorescence` is set to `NaN`.

B

Appended papers

1. J. Uhlendorf, S. Bottani, F. Fages, P. Hersen, and G. Batt. Towards real-time control of gene expression: controlling the HOG signaling cascade. *Pacific Symposium On Biocomputing*, 338–349, 2011.

2. J. Uhlendorf, P. Hersen, and G. Batt. Towards real-time control of gene expression: *in silico* analysis. *The 18th World Congress of the International Federation of Automatic Control (IFAC)*, 2011.

3. J. Uhlendorf, A. Miermont, T. Delaveau, G. Charvin, F. Fages, S. Bottani, G. Batt, and P. Hersen. Long-term model predictive control of gene expression at the population and single-cell levels. *Proceedings of the National Academy of Sciences*, 109(35):14271–14276, 2012.

4. A. Miermont, J. Uhlendorf, M. McClean, P. Hersen, et al. The dynamical systems properties of the HOG signaling cascade. *Journal of signal transduction*, 2011:930940, 2011.

B.1 Appended paper 1: Towards real-time control of gene expression: controlling the HOG signaling cascade

Towards real-time control of gene expression: controlling the HOG signaling cascade

Jannis Uhlendorf, Samuel Bottani, François Fages, Pascal Hersen and Gregory Batt

Pacific Symposium on Biocomputing, 16:338-349, 2011

TOWARDS REAL-TIME CONTROL OF GENE EXPRESSION: CONTROLLING THE HOG SIGNALING CASCADE

JANNIS UHLENDORF[1,2], SAMUEL BOTTANI[2], FRANÇOIS FAGES[1], PASCAL HERSEN[2‖], GREGORY BATT[1♯]

[1] *Contraintes research group, Institut National de Recherche en Informatique et en Automatique, INRIA Paris–Rocquencourt, France*
♯*contact: gregory.batt@inria.fr*

[2] *Laboratoire Matière et Systèmes Complexes, Université Paris Diderot and Centre National de la Recherche Scientifique, UMR 7057, Paris, France*
‖*contact: pascal.hersen@univ-paris-diderot.fr*

To decipher the dynamical functioning of cellular processes, the method of choice is to observe the time response of cells subjected to well controlled perturbations in time and amplitude. Efficient methods, based on molecular biology, are available to monitor quantitatively and dynamically many cellular processes. In contrast, it is still a challenge to perturb cellular processes - such as gene expression - in a precise and controlled manner. Here, we propose a first step towards *in vivo* control of gene expression: in real-time, we dynamically control the activity of a yeast signaling cascade thanks to an experimental platform combining a micro-fluidic device, an epi-fluorescence microscope and software implementing control approaches. We experimentally demonstrate the feasibility of this approach, and we investigate computationally some possible improvements of our control strategy using a model of the yeast osmo-adaptation response fitted to our data.

1. Introduction

To understand biology at the system level, one has to study both the *structure* and the *dynamics* of cellular processes [18,19,32]. On the one hand, genetic analyses are required to analyze the structure of signaling pathways and genetic networks. On the other hand, to access to the dynamical functioning of cellular processes, one has to observe the time response of cells to well controlled perturbations. Hence, the information level provided by experiments crucially depends on our capacity to observe *and* perturb biological systems at the cellular level. Efficient experimental tools have been developed to *monitor* both quantitatively and dynamically many cellular processes. Gene expression can be measured through micro-arrays or quantitative RT-PCR and conveniently observed at the single cell level through the combination of fluorescent reporter proteins and FACS techniques or microscopy [19,21,26]. In contrast, it is still a challenge to *perturb* cellular processes in a precise and controlled manner. A commonly used strategy resides on using inducible promoters to modulate the expression of a gene of interest by the addition of a diffusible molecule in the external cellular environment [12, 16,28]. However, even if the activity of the inducible promoter can be modulated quantitatively, there is no guarantee that the target gene will reach a desired constant expression level over a long period of time. Indeed, variations may arise because of modifications of the physiological state of the cell due to internal feedback loops and cellular adaptation. The expression of a transcription factor regulating itself is even more problematic. Moreover, both theoretical [1, 15] and recent experimental [14,24] results demonstrate the need for elaborate, time-varying

perturbations to decipher quantitatively certain dynamics features of cellular responses. This notably includes the numerous biological processes in which the timing of gene expression plays a central role such as the regulation of the cell cycle. To summarize, existing solutions for the artificial control of gene expression are dissatisfying on two counts since, (i) expressing a gene of interest in a well-controlled, sustained way cannot be conveniently realized at the present time, and (ii) the investigation of certain dynamical properties necessitates dynamical, time-varying perturbations of gene expression for which no solution is currently available.

Here, we propose a first step towards *in vivo* control of gene expression. We have implemented an experimental platform for the *in vivo* control of a signaling pathway in *Saccharomyces cerevisiae*. We chose to control the activity of the HOG cascade which is activated in response to hyper-osmotic perturbations and promotes the transcription of osmo-adaptative response genes. Given a desired temporal profile, the activity of the signaling cascade is monitored in real time and deviations from the desired values are dynamically corrected by varying the osmolarity of the cellular environment (Fig. 1). This can be achieved thanks to a dedicated micro-fluidic device. This experimental platform is driven by software, that notably implements control algorithms, responsible for computing how the cellular environment (osmolarity) should be modified to correct the observed deviations from target values.

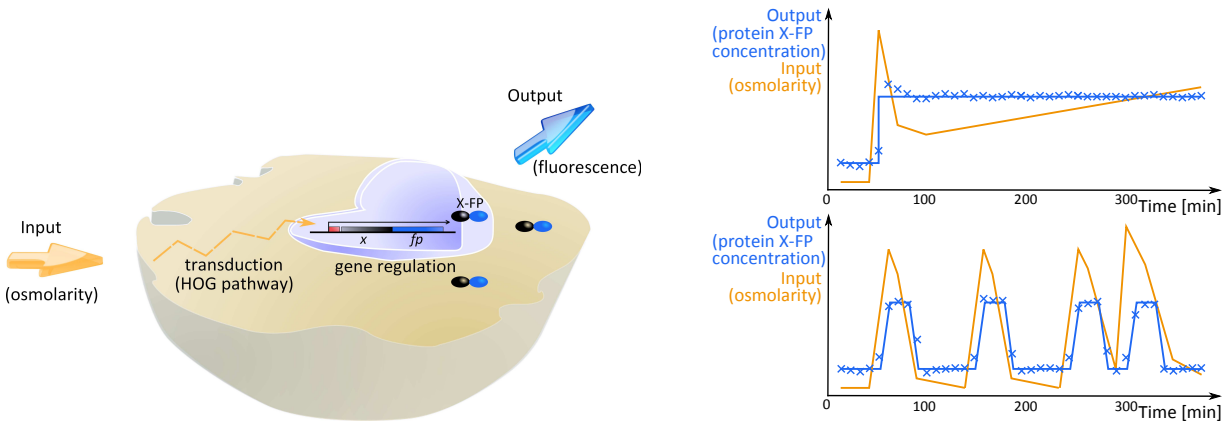


Fig. 1: The control problem. (a) Schematic Input/Output description of the cell. (b) Schematic representation of a desired output (blue), an applied input (orange) and the obtained output (blue crosses), for two different situations. In the first case (top), the goal is to dynamically maintain the concentration of a target protein X, fused to a fluorescent protein (FP), at a constant level. In the second case (bottom), the goal is to create a complex perturbation signal by varying the concentration of the protein X with time.

The work presented here differs significantly from previous applications of control theory in systems and synthetic biology contexts. So far, control theory contributions consisted essentially to shed a new light on biological phenomena, notably by suggesting underlying organization principles in biology [8,9]. An illustrative example is the use of the notion of integral feedback control to explain the robust perfect adaptation observed in bacterial chemotaxis [34]. Other insightful examples are given in a recent textbook [15]. Control theory has also been used in optimal experimental design applications [1,23]. But quite surprizingly, only a few

(theoretical) studies focused on the actual control on a biomolecular process, e.g. [2,4,7,17]. Moreover, to the best of our knowledge, control theory has not yet been applied *in vivo* for the *actual feedback control of biological cellular processes at the single cell level*.

The paper is organized as follows. In section 2, we present in details the proposed platform for real-time control of the HOG signaling cascade activity and gene expression. In Section 3, we present preliminary experimental results obtained when controlling the nuclear localization of the Hog1 protein. This represents an essential first step towards controlling gene expression. In Section 4, we discuss possible improvements of our control approach using a simple model of the osmotic stress response. Conclusions are provided in the last section.

2. A platform for real-time control of gene expression

2.1. *An integrated real-time control platform*

Central to control theory is the notion of *feedback control* [30]. The idea is to compute the inputs to apply at the next time instant in function of outputs previously obtained. This way, knowledge of past errors is used to improve the control. In comparison to open loop control where the control strategy is computed beforehand, closed loop control approaches are generally less sensitive to model uncertainties and can compensate for external disturbances. These two features are highly desirable for any biological application. However, performing a control in real-time necessitates a tight integration between measurement device, control software and actuator.

As described in Figure 2, the HOG pathway activity can be monitored at the single cell level using time lapse fluorescent microscopy. The cellular environment can be controlled using the micro-fluidic device developed by Hersen and colleagues [14]. Not only this device allows a fast and well-controlled change of the cellular environment, but also, it guarantees that with the exception of the input signal the cell environment is otherwise held constant. We implemented algorithms for image analysis, state estimation and input computation in a Matlab program that communicates with and drives the microscope via MicroManager [31] and the micro-fluidic pressure controller.

2.2. *Using HOG signaling cascade*

To link external environmental changes to gene expression, we use a natural signaling pathway: the Hyper Osmolar Glycerol (HOG) pathway in the yeast *Saccharomyces cerevisiae*. This MAP kinase pathway is used to sense osmolar pressure changes in the environment and to trigger osmotic stress responses that maintain water homeostasis [29]. More precisely, two osmo-sensor proteins (Sln1 and Sho1) transduce the signal to the Hog1 protein via a phosphorylation cascade. Once phosphorylated, Hog1 promotes the osmo-adaptative response in at least three different ways. Firstly, Hog1 translocates into the nucleus and alters, directly or indirectly, the expression of a large number of genes [27]. Secondly, Hog1 has also a cytoplasmic activity since it regulates negatively glycerol export by inhibiting the activity of the Fps1 glycerol channel [3]. Thirdly, Hog1 activates glycerol producing enzymes, notably Gpd1 [33]. Hence, the osmo-adaptative response involves at least three natural feedback loops.

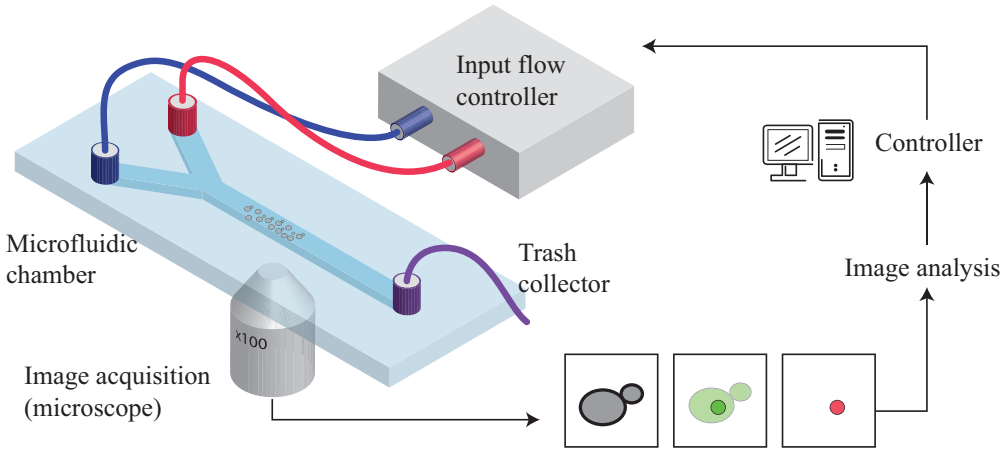


Fig. 2: The integrated control platform. The main elements of the feedback loop are (i) a microfluidic device allowing a rapid control of the cellular environment, (ii) a microscope for phase contrast and fluorescence measurements, (iii) yeast cells with Hog1, a nuclear marker (Htb2) and the protein of interest (X) fused to compatible fluorescent markers, and (iv) Matlab software for image analysis and controller implementation.

Our motivation for using this pathway is triple. Firstly, it has been extensively experimentally studied and quantitative models are available [6,14,20,22,24,25,35]. Second, the output of the signal transduction pathway can be experimentally quantified. Indeed, if Hog1 is fused to a fluorescent protein, its nuclear localization can be quantified and provides a measure of the Hog1 activity [10]. Thirdly, it has been experimentally shown that for fast osmolarity changes, the pathway integrates the signal: the transduction pathway acts as a low-pass filter with a bandwidth approximatively equal to 5×10^{-3} Hz [14]. This property allows us to *emulate* an analog control by rapidly switching (frequencies greater than 0.1 Hz) between two media: the normal growth medium and a sorbitol enriched ($\sim 1\,M$) medium. For example, a two minute osmotic stress corresponding to a $0.4\,M$ sorbitol intensity is obtained by flowing cells 12 times with normal medium during $6\,s$. and with sorbitol-rich medium during $4\,s$.

In this paper, we use a yeast strain with Hog1 fused to GFP and the nuclear protein Htb2 fused to mCherry [14]. The latter is used to conveniently localize the nuclear region. We define the relative Hog1 nuclear localization $h(t)$ as the ratio of the mean fluorescence pixel intensities of Hog1-GFP in the nucleus and in the cytoplasm.

$$h(t) = \frac{\langle Pixel\ intensity \rangle_{nuc}}{\langle Pixel\ intensity \rangle_{cyto}}$$

The normalized Hog1 nuclear localization $h_n(t)$ is then simply $h_n(t) = h(t)/h(t_0)$. These definitions are motivated by the fact that this gives measures that are relatively robust with respect to fluorescent protein photo-bleaching and cell-to-cell variations.

3. Controlling transcription factor nuclear localization using a simple control approach

In this section, we present preliminary results obtained on controlling the Hog1 nuclear localization. The control of Hog1 nuclear activity is a prerequisite for utilizing the Hog pathway

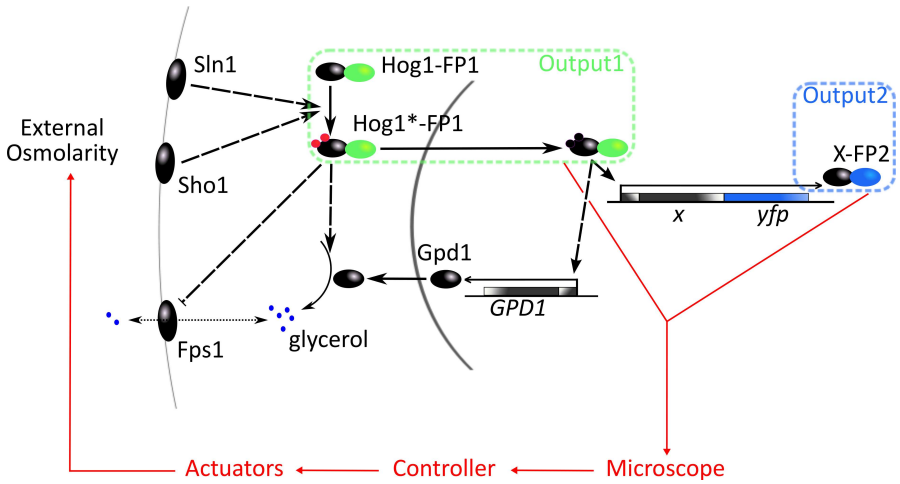


Fig. 3: Schematic representation of the HOG pathway with natural and engineered feedbacks. Solid and dashed arrows indicate direct and indirect effects, respectively. For a detailed description, see the main text. FP1 and FP2 in the figure denote two different fluorescent proteins.

to control gene expression. As a matter of fact, controlling the duration of activated Hog1 residence in the nucleus will lead to bursts of expression for the genes which are placed under Hog1 dependent promoter. There are different options how to encode a certain gene expression profile. We could either work with a constantly high signal and adjust the amount of Hog1 in the nucleus (amplitude modulation), or we could successively activate the Hog pathway for a short duration and control the frequency of these activations (frequency modulation). To test these two alternative strategies, we consider two problems: maintaining a given constant level of Hog1 nuclear localization over a long time period, or obtaining pulses of Hog1 nuclear localization in a repeated manner. These results have been obtained using the simplest control approach: a PID controller.

3.1. *PID control*

A proportional-integral-derivative (PID) controller is a generic closed-loop control algorithm, generic meaning that is does not require any structural knowledge about the controlled system [30]. Due to its simplicity this type of control is very often applied in engineering applications. A PID controller measures the deviations ("errors") of measured states from target states, and uses this information to compute the control. The applied control u at time t is the weighted sum of the error, $e(t)$, its derivative and the integral of past errors $e(\tau)$, $\tau \in [0, t]$:

$$u(t) = k_p \cdot e(t) + k_i \cdot \int_0^t e(\tau)d\tau + k_d \cdot \frac{d}{dt}e(t)$$

where k_p, k_i and k_d are the proportional, integral and derivative gains.

In our case the error $e(t)$ is the difference between the measured normalized Hog1 nuclear localization $h_n(t)$ and its reference value at the corresponding time point. Because we consider tracking problems, only the recent past errors are relevant. Therefore, we integrate the error only on the interval $[t-\Delta, t]$, where Δ is approximatively 2 minutes. We tuned the controller gains manually using a trial and error approach. The derivative term, and to a lesser extend,

the proportional term are responsible for implementing a fast system response to target value changes. However large values for these parameters favor oscillations and loss of stability. In practice, we found that setting the derivative gain to zero and using values for k_p and k_i close to 2 and 1.5 leads to a good compromise between response time and stability in our experimental setting.

3.2. *Experimental results*

We designed two control experiments to test the possible strategies discussed above: using amplitude or frequency encoding. The first type of experiment is to try to maintain the system output at a constant target level (Fig 4 left). Quantitatively, the relative Hog1 nuclear localization should remain 20% higher than its nominal value in unstressed cells. The second type of experiment is to try to obtain repeatedly trapezoidal motifs. The amplitude of output variations also corresponds to a 20% increase above nominal value (Fig 4 right).

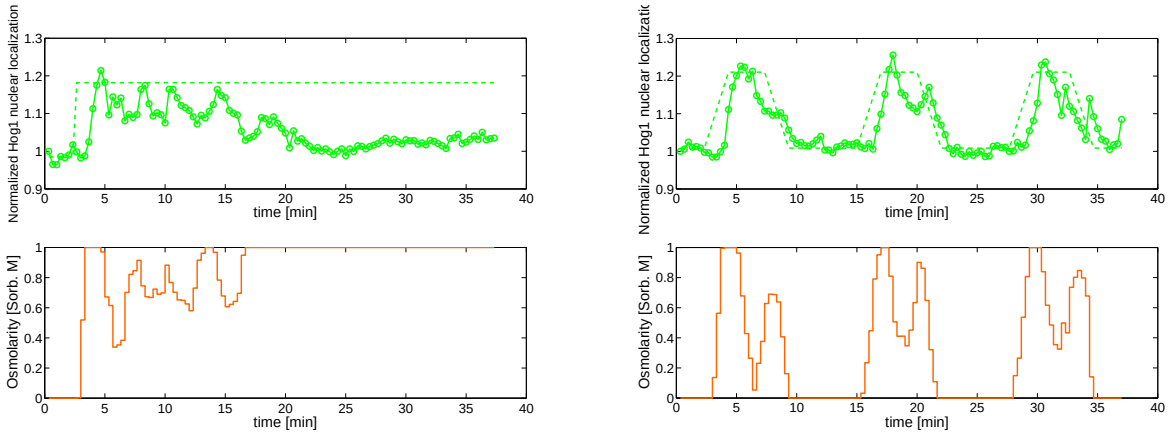


Fig. 4: Experimental results for the control of Hog1 nuclear localization. Left: Controlling the amplitude of Hog1 localization does only work for short durations due to internal feedback and cellular adaptation to sustained hyper osmotic conditions. Right: With a frequency encoded signal, the cell is able to reset between successive shocks and follows the reference values for the whole experiment.

Our experimental results clearly show that the control is effective in yeast cells. Consider for example the step experiment at time 2, when the target value changes. Following this change, the controller applies an osmotic stress, resulting after a 1-2 minute delay in an increase in the Hog1 nuclear localization. Then the system overshoots and the controller decreases the osmolarity of the environment. Oscillations ensue around a level below the target value during approximatively 15 minutes, during which increasing inputs are applied. Finally, even the maximal input is not sufficient to prevent the system from drifting away towards its nominal level.

The interpretation of these control results is simple. Because of internal natural feedbacks, the cells adapt (notably produce glycerol) and become insensitive to high osmolarity environments. Therefore, unless all internal feedback loops are inactivated, the amplitude-based control strategy seems not feasible. The inability of the controller to maintain the output at the

target value in osmo-sensitive cells can be explained by the initiation of an osmo-adaptative response causing cells to drift away from the target value, together with the use of a rather narrow integration window in our PID controller.

Concerning the repeated motif experiment, it is fair to say that despite time lags and a relatively noisy behavior, the controller succeeds in producing the desired time varying output (Fig 4 right). As it appears on the plots, the 6 minute time separation between the 8 minute long motifs seem sufficient to fully reset the system to its normal, osmo-sensitive state. Based on these experimental results, the frequency encoding strategy for gene expression seems promising. However, before dealing with the actual control of gene expression, improvements in our control approach are needed. The capacity of the controller to predict rather than just to react -this would help dealing with the lag problem-, and the capacity to filter noise out -this would make the control more robust- are two features of significant interest.

4. Design of an improved control approach

The major advantage of the PID controller is that it does not rely on a model of the system. This makes it particularly easy to deploy. However, performances achieved using model-based control approaches are generally superior. In this section, we use the simple model proposed by Muzzey and colleagues [25], fitted to our data, to compare performances obtained with the PID controller and a model based control approach.

4.1. *Development of a simple linear model*

Numerous models have been developed for the osmotic stress response [6,14,20,22,24,25,35]. Because of its capacity to capture essential aspects of the HOG pathway, including notably the cell adaptation, and of its mathematical simplicity, we reuse the three variables linear model developed by Muzzey and colleagues [25]. In short, the state of the system is described by three variables, s_1, s_2, and s_3, corresponding to the nuclear Hog1 enrichment, its time integral, and glycerol relative concentration, respectively, and one input, u, corresponding to the external osmolarity. Since the osmo-stresses studied in [25] are caused by a different osmolyte (salt versus sorbitol), we introduce a factor σ to rescale the input u, if needed.

$$\dot{s}_1 = k_h\,(\sigma\,u - s_3) - \gamma_h\,s_1 \tag{1}$$

$$\dot{s}_2 = \alpha_d\,s_1 \tag{2}$$

$$\dot{s}_3 = s_2 + \alpha_i\,(\sigma\,u - s_3) - \gamma_g\,s_3 \tag{3}$$

In the above model, $\sigma\,u - s_3$ corresponds to the net osmolarity effectively sensed by the cell. In hyper-osmotic conditions, the production of intracellular glycerol (s_3) and the Hog1 nuclear localization (s_1) are increased. The increased Hog1 nuclear localization increases s_2 and hence s_3. Therefore one distinguishes a *direct* and an *indirect* effect of hyper-osmotic stress on glycerol accumulation [25].

To fit the model parameters to our system we perform two types of experiments in which cells are exposed to hyper-osmotic stresses differing either in magnitude or duration. The first set of experiments is used primarily to estimate the relation between osmotic stress and Hog1 localization, whereas the second set of experiments is used primarily to investigate the cell

dynamical adaptation to osmotic stress. One should note that we experimentally measure the normalized Hog1 nuclear *localization* $h_n(t)$, whereas the variable $s_1(t)$ in the Muzzey model corresponds to the Hog1 nuclear *enrichment*. However, it holds that $h_n(t) = s_1(t) + 1$ [25]. In the sequel, to allow for comparison with the experimental results of Section 3, we present all our results -experimental and computational- using $h_n(t)$.

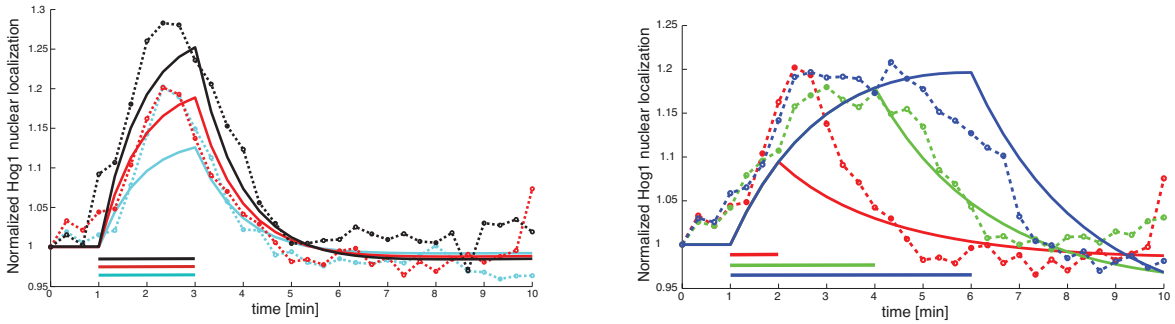


Fig. 5: Cell response to different hyper-osmotic stresses. Left: Stresses of different intensities. Cyan, red, and black plots correspond to a 0.4, 0.6, or 0.8 M stress applied during 2 minutes. Right: Stresses of different durations. Red, green and blue plots correspond to 0.6 M stress applied during 2, 4, or 6 minutes. Dashed and solid lines represent experimental data and model predictions, respectively.

To find parameter values for our model, we use the state-of-the-art non-linear optimization tool CMAES implementing a covariance matrix adaptation evolution strategy [13]. The objective function to minimize is the sum of a mean square error term, where the error is the difference between measured and predicted values for s_1, and a penalty term enforcing the positiveness of parameters and initial conditions. Parameter estimates are then manually fine-tuned (see Table 1).

k_h	γ_h	γ_g	α_h	α_g	σ
1.984	0.9225	0.5950	0.1612	0.0106	0.2

Table 1: Parameter values fitted to the experimental data shown in Fig. 5. All parameter units are min^{-1}, excepted for the dimensionless parameter σ.

As can be seen from the plots shown in Fig. 5, the model is able to capture qualitatively, and up to some degree, quantitatively, the behavior of yeast cells subjected to hyper-osmotic stresses. This is commendable given the extreme simplicity of the model.

4.2. *Comparison of different control approaches*

Equipped with a model of our system, we can computationally simulate the system response and compare various control approaches. Given the time-consuming aspect of experiments, working on simulated but realistic data allows us to rapidly test alternative control approaches. When computationally testing a model-based control approach, one uses the same model in the simulator and in the model based controller. That is, the model based controller knows

perfectly the system dynamics. To make fair comparisons, we assume that only the output (and not the full state) is visible by the controller and we add (Gaussian) noise to the system output.

We present here a model predictive control (MPC) approach. The objective of MPC is to minimize the difference between the simulated and the target outputs by using a receding horizon strategy: given an estimate of the current state of the system, a control strategy to be applied during a short time horizon is searched for, and applied for a short period of time. Then, the approach is applied again, with the estimation of the new state, and the computation of a control strategy for a new short time horizon. This receding horizon strategy yields an effective feedback control [11]. Because MPC applies to linear and nonlinear systems, this approach can easily be extended to deal with future improved models. An other motivation for using MPC rather than the conventional control approach for linear system output tracking, based on a linear quadratic gaussian controller [30], is that simple non-linear constraints (e.g. bounded input) can easily be integrated in this framework.

For our application, we implement an MPC approach using Kalman filtering and a simple search strategy. The use of a Kalman filter is a standard approach to estimate the full state of a linear system based on (noisy) observations [30]. Then, at time t, we search for three input values, u_1, u_2, and u_3, that when applied on the time intervals $[t, t+1]$, $]t+1, t+2]$, and $]t+2, t+3]$, respectively, minimize the squared error, again defined as the distance between the target and the simulated outputs, on the time interval $[t, t+3]$. u_1 is applied on $[t, t+1]$ and the procedure is restarted at time $t+1$. At each iteration, we use CMAES, a global optimization, tool to search for the three input osmolarities u_1, u_2, and u_3. Naturally, in our setting, the input (osmolarity) is necessarily positive and bounded. Therefore, we limit the search to the interval of feasible osmolarities. The computational effort remains limited, since less than one second is needed for each iteration. For comparison, the timestep duration of the control loop in our experiments is $20\,s$. So using MPC does not challenges the real-time requirement.

We also consider here the PID controller presented in Section 3, but applied on simulated data as explained in this section. All these computational procedures have been implemented in Matlab.

The results obtained with the two control strategies and the for two different control problems are shown in Figure 6. Regarding the difficulty to maintain pathway activity over a prolonged period and the feasibility of creating repeated short time activity patterns, the results obtained with both control approaches are fully consistent with our experimental findings. The comparison of the results obtained with the PID controller on the experimental (Fig 4) and simulated data (Fig 6) shows that the PID performs better in the second case. This might be explained by a higher complexity of cellular variability (ie the "noise" is not just plain Gaussian). As expected, the lag and incomplete drift compensation are also observed on simulated data, albeit attenuated. In contrast, the model predictive results show neither. This corroborates the fact that they originate -at least partly- from the reactive rather than predictive nature of the PID controller. Moreover, the control is also much more regular in the MPC experiments. Very likely, this comes from the use of Kalman filtering. One should note that this is not due to an improper parametrization of the PID. Indeed the relatively

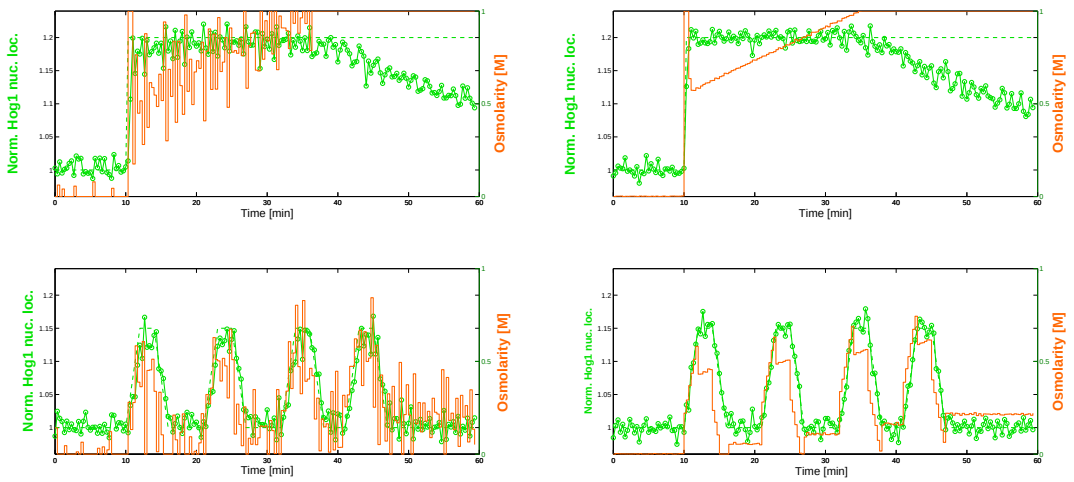


Fig. 6: Comparison of PID (left) and MPC (right) control strategies for two different control problems and on simulated data. *Norm. Hog1 nuc. loc.* stands for normalized Hog1 nuclear localization.

high proportional gain that causes large input changes is needed to ensure a fast response.

To summarize, the model predictive control approach is superior on all counts to the PID controller, at the cost of a very limited computational overhead. However, one should stress that the quality of a model based controller ultimately depends on the quality of the model of the system. So to effectively apply MPC on yeast cells, significant modeling work might be needed. But then one will have the effective proof that the main features of the osmo-adaptative response are captured in sufficient details.

5. Discussion

We presented an integrated experimental platform and demonstrated the feasibility of controlling the nuclear localization of the protein Hog1. Stated differently, we have shown how to create a dynamically controlled inducible promoter. As a matter of fact, it should be possible to place any gene under the control of a Hog1-dependent promoter and then to force its expression by controlling Hog1 nuclear localization. Consequently, this contribution describes a first, crucial step towards real-time control of gene expression.

Using the HOG pathway has several advantages, the most important ones being its quick activation and de-activation which are crucial to ensure efficient dynamics of the control loop, and the established correlation between nuclear localization and activity. It is to be noted though, that contrarily to known inducible promoters such as the Tet system, activating the HOG pathway also affects the cell physiological state, since many genes are transcribed to ensure proper cellular response to an hyper osmotic environment. For real applications, one should achieve a clear separation between controlling gene expression dynamically and altering the physiological state. This might require engineering the HOG cascade, or using other alternative signaling pathways with similar dynamics and nuclear translocation.

Interestingly, our results suggest that for our application, it is preferable to use *frequency*

encoding to control gene expression. Indeed, because of fast, non-genetic adaptation feedbacks, the output of the signaling cascade can not be held constant over a prolonged period. A Frequency encoding strategy is widely used by neural networks which computing activity relies on action potential pulses. Although it is generally assumed that gene regulation is naturally controlled by amplitude modulation, a recent study by Elowitz's team showed that the expression of some genes in yeast are regulated by the frequency of expression bursts led by the transcription factor Crz1 [5]. The authors proposed that the functional role of frequency modulation is to ease the coordination of the expression of multiple target genes. Based on our results, one can propose an alternative role of regulation by frequency modulation: it allows for both a rapid non-genetic response and a slower transcriptional response leading to a complete adaptation to a given stress.

In a future work, we will use a model-based control approach to improve our results on Hog1 nuclear localization. Moreover, we will progress towards our main goal, that is, gene expression control, by studying a candidate gene fused to a fluorescent tag under the direct control of Hog1. The control platform will be adapted to read as outputs both the localization of Hog1 and the actual expression level of the gene of interest.

We anticipate that this platform to tune in real-time the level of expression of a gene of interest will be a useful tool for the biologist to better understand living processes in single cells. Quoting Feynman saying 'what I cannot built, I cannot understand', synthetic biologists propose that building systems helps to better understand them. Here, we propose that controlling them is an effective way to assess our understanding: what I cannot control, I have not understood.

References

1. J.F. Apgar, J.E. Toettcher, D. Endy, F.M. White, and B. Tidor. Stimulus design for model selection and validation in cell signaling. *PLoS Computational Biology*, 4(2):e30, 2008.
2. S. Azuma, E. Yanagisawa, and J. Imura. Controllability analysis of biosystems based on piecewise affine systems approach. *IEEE Transactions on Circuits and Systems and IEEE Transactions on Automatic Control*, 53, 2008. Joint special issue on Systems Biology.
3. S. E. Beese, T. Negishi, and D.E. Levin. Identification of positive regulators of the yeast Fps1 glycerol channel. *PLoS Genetics*, 5(11):e1000738, 2009.
4. C. Belta, L.C.G.J.M. Habets, and V. Kumar. Control of multi-affine systems on rectangles with applications to hybrid biomolecular networks. In *Proceedings of the 41st IEEE Conference on Decision and Control, CDC'02*, 2002.
5. L. Cai, C.K. Dalal, and M.B. Elowitz. Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature*, 455(7212):485–490, 2008.
6. A.P. Capaldi, T. Kaplan, Y. Liu, N. Habib, A. Regev, N. Friedman, and E. K O'Shea. Structure and function of a transcriptional network activated by the MAPK Hog1. *Nature Genetics*, 40(11):1300–1306, 2008.
7. M. Chaves and J.-L. Gouzé. Exact control of genetic networks in a qualitative framework: the bistable switch example. Technical Report RR-7359, INRIA Sophia-Antipolis, 2010.
8. M.E. Csete and J.C. Doyle. Reverse engineering of biological complexity. *Science*, 295(5560):1664–1669, 2002.
9. H. El-Samad, H. Kurata, J.C. Doyle, C.A. Gross, and M. Khammash. Surviving heat shock: Control strategies for robustness and performance. *Proceedings of the National Academy of Sciences of the USA*, 102(8):2736–2741, 2005.
10. P. Ferrigno, F. Posas, D. Koepp, H. Saito, and P.A. Silver. Regulated nucleo/cytoplasmic exchange of

HOG1 MAPK requires the importin beta homologs NMD5 and XPO1. *EMBO Journal*, 17(19):5606–5614, 1998.

11. R. Findeisen, F. Allögwer, and L. Biegler. *Assessment and Future Directions of Nonlinear Model Predictive Control*, volume 358 of *LNCIS*. Springer, 2007.
12. M. Gossen, S. Freundlieb, G. Bender, G. Muller, W. Hillen, and H. Bujard. Transcriptional activation by tetracyclines in mammalian cells. *Science*, 268(5218):1766–1769, 1995.
13. N. Hansen and A. Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195, 2001.
14. P. Hersen, M. N. McClean, L. Mahadevan, and S. Ramanathan. Signal processing by the HOG MAP kinase pathway. *Proceedings of the National Academy of Sciences of the USA*, 105(20):7165–7170, 2008.
15. P.A. Iglesias and B.P. Ingalls. *Control Theory and Systems Biology*. MIT Press, 2009.
16. F. Jacob and J. Monod. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, 3:318–356, 1961.
17. A.A. Julius, A. Halasz, M.S. Sakar, H. Rubin, V. Kumar, and G.J. Pappas. Stochastic modeling and control of biological systems: the lactose regulation system of *Escherichia coli*. *IEEE Transactions on Circuits and Systems and IEEE Transactions on Automatic Control*, 53, 2008. Joint special issue on Systems Biology.
18. H. Kitano. Systems biology: A brief overview. *Science*, 295(5560):1662–1664, 2002.
19. E. Klipp, R. Herwig, A. Kowald, C. Wierling, and H. Lehrach. *Systems biology in practice: Concepts, implementation and application*. Wiley Press, 2005.
20. E. Klipp, B. Nordlander, R. Kruger, P. Gennemark, and S. Hohmann. Integrative model of the response of yeast to osmotic shock. *Nature Biotechnology*, 23(8):975–982, 2005.
21. K.M. Klucher, M.J. Gerlach, and G.Q. Daley. A novel method to isolate cells with conditional gene expression using fluorescence activated cell sorting. *Nucleic Acids Research*, 25(23):4858–4860, 1997.
22. J. Macia, S. Regot, T. Peeters, N. Conde, R. Solé, and F. Posas. Dynamic signaling in the Hog1 MAPK pathway relies on high basal signal transduction. *Science Signaling*, 2(63):ra13, 2009.
23. F. Menolascina, D. Bellomo, T. Maiwald, V. Bevilacqua, C. Ciminelli, A. Paradiso, and S. Tommasi. Developing optimal input design strategies in cancer systems biology with applications to microfluidic device engineering. *BMC Bioinformatics*, 10(Suppl 12):S4, 2009.
24. J.T. Mettetal, D. Muzzey, C. Gomez-Uribe, and A. van Oudenaarden. The frequency dependence of osmo-adaptation in Saccharomyces cerevisiae. *Science*, 319(5862):482–484, 2008.
25. D. Muzzey, C.A. Gomez-Uribe, J.T. Mettetal, and A. van Oudenaarden. A systems-level analysis of perfect adaptation in yeast osmoregulation. *Cell*, 138(1):160–171, 2009.
26. D. Muzzey and A. van Oudenaarden. Quantitative time-lapse fluorescence microscopy in single cells. *Annual Review of Cell and Developmental Biology*, 25(1):301–327, 2009.
27. S.M. O'Rourke and I. Herskowitz. Unique and redundant roles for HOG MAPK pathway components as revealed by whole-genome expression analysis. *Molecular Biology of the Cell*, 15(2):532–542, 2004.
28. T.M. Roberts, R. Kacich, and M. Ptashne. A general method for maximizing the expression of a cloned gene. *Proceedings of the National Academy of Sciences of the USA*, 76(2):760–764, 1979.
29. W.H. Mager S. Hohmann. *Yeast stress responses*. Topics in Current Genetics. Springer, 2003.
30. E.D. Sontag. *Mathematical Control Theory. Deterministic Finite-Dimensional Systems*. Springer, 1998.
31. N. Stuurman and A. Edelstein. μmanager: the open source microscopy software v.1.3, 2010. http://www.micro-manager.org.
32. Z. Szallasi, J. Stelling, and V. Periwal, editors. *System Modeling in Cellular Biology: From Concepts to Nuts and Bolts*. MIT Press, 2006.
33. P.J. Westfall, J.C. Patterson, R.E. Chen, and J. Thorner. Stress resistance and signal fidelity independent of nuclear MAPK function. *Proceedings of the National Academy of Sciences of the USA*, 105(34):12212–12217, 2008.
34. T.-M. Yi, Y. Huang, M.I. Simon, and J.C. Doyle. Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proceedings of the National Academy of Sciences of the USA*, 97(9):4649–4653, 2000.
35. Z. Zi, W. Liebermeister, and E. Klipp. A quantitative study of the Hog1 MAPK response to fluctuating osmotic stress in Saccharomyces cerevisiae. *PLoS ONE*, 5(3):e9522, 2010.

Towards real-time control of gene expression: in silico analysis

Jannis Uhlendorf, Pascal Hersen and Gregory Batt

The 18th World Congress of the International Federation of Automatic Control (IFAC), 18:14844-14850, 2011

Preprints of the 18th IFAC World Congress
Milano (Italy) August 28 - September 2, 2011

Towards Real-Time Control of Gene Expression: *in silico* Analysis ⋆

Jannis Uhlendorf [1,2] **Pascal Hersen** [2] **Gregory Batt** [1,*]

[1] *INRIA Paris-Rocquencourt, F-78153 Le Chesnay, France*
[2] *Matière et Système Complexes Lab, UMR CNRS 7057, 10 rue A. Domon et L. Duquet, F-75205 Paris cedex 13, France*
[*] *contact: gregory.batt@inria.fr*

Abstract: One major goal of systems biology is to understand the dynamical functioning of biological systems at the cellular level. A common approach to investigate the dynamics of a system is to observe its response to perturbations. To improve our capacity to perturb cellular processes via the expression of a given protein with a chosen temporal expression profile, we develop an experimental platform for the real time control of gene expression. In short, this platform allows for applying short osmotic stresses to yeast cells, that trigger the expression of a target gene via the activation of the HOG signal transduction pathway, and for observing in real time the cellular response. In Uhlendorf et al. (PSB'11), we describe preliminary experimental results on the control of the signal transduction pathway obtained using a simple proportional-integral controller. However, the control of the full system, including the much slower transcription and translation processes, necessitates more elaborate control methods.
In this paper, we propose a model based control strategy tailored to the specificities of the biological system, notably to its perfect adaptation to osmotic stress. The practical feasibility and the robustness of the proposed approach with respect to gene expression noise is tested *in silico* using a simple, switched linear model of the osmostress response in yeast.

Keywords: Kinetic modeling and control of biological systems; control of gene expression; model predictive control; switched linear system; yeast hyperosmotic stress response

1. INTRODUCTION

One major goal of systems biology is to understand the dynamical functioning of biological systems at the cellular level. One common approach to investigate the dynamics of a system is to observe its response to perturbations. For the development of a quantitative understanding (and of quantitative models), precision in the observation *and* in the perturbation are both important. With the development of fluorescent markers, our capacity to observe cellular processes at the single cell level has recently greatly improved (Muzzey and van Oudenaarden, 2009). In contrast, our capacity to perturb cellular processes in a well-controlled, quantitative manner remains very limited. This is particularly striking if one considers time-varying perturbations, that are however highly informative on system's dynamics. This can notably be explained by the large variability of individual cell responses to a same external stimulation (Colman-Lerner et al., 2005; Newman et al., 2006).

To improve our capacity to perturb cellular processes via the expression of a given protein with a chosen temporal expression profile, we develop a platform for the real time control of gene expression (Uhlendorf et al., 2011). In short, this platform allows for applying short osmotic stresses to yeast cells, that trigger gene expression via the activation of the so-called Hyper Osmotic Glycerol (HOG) signal transduction pathway, and for observing in real time the cellular response. In Uhlendorf et al. (2011), we describe preliminary experimental results on the control of the activity of the signal transduction pathway, obtained using a simple proportional-integral (PI) controller. However, the control of the full system, being more complex and having a significant inertia, will likely require more elaborate control methods.

In this paper, we propose a control strategy tailored to the specificities of our biological problem. In particular, we decompose the system into two subsystems, corresponding to signal transduction and gene expression, and having significantly different response times. We exploit the 'cascaded' structure of the system to propose a two-layer model predictive control strategy. Additionally, to skirt the rapid cell adaptation to stress, we adopt a pulse-modulated strategy to control gene expression. The practical feasibility and the robustness of the proposed approach with respect to noise in the gene expression is tested *in silico* using a simple, switched linear model of the osmostress response in yeast.

The paper is organized as follows. In section 2, we briefly present the control platform and the controlled system. This brief description of the experimental setup motivates important modeling and control assumptions made in the

⋆ This work is partially supported by the INRIA/INSERM Colage Action d'Envergure

following section. In section 3, we detail the proposed control strategy. Then, we test its feasibility and robustness with respect to biological variability on two simple control problems. The last section summarizes our work and discusses results in the context of related work.

2. THE CONTROLLED SYSTEM AND THE CONTROL PLATFORM

The overall objective is to place the expression of a protein under the control of an external signal and to control this expression in a precise temporal manner. This necessitates signal transduction and gene expression. We have chosen to exploit a natural signal transduction and gene expression pathway: the High Osmolarity Glycerol (HOG) pathway in the yeast *Saccharomyces cerevisiae*.

2.1 Exploiting the natural osmostress response

The HOG pathway senses osmolar pressure changes in the environment. The activation of this pathway orchestrates cellular osmotic stress responses that maintain water homeostasis (Hohmann and Mager, 2003). More precisely, two osmo-sensor proteins (Sln1 and Sho1) transduce the signal to the Hog1 protein via a phosphorylation cascade. Once phosphorylated, Hog1 triggers different osmo-adaptative responses that essentially favor the accumulation of glycerol in the cell, thus restoring the osmotic balance of the cell with its environment. Firstly, Hog1 regulates negatively Fps1 glycerol channels, thereby preventing glycerol export (Beese et al., 2009). Secondly, Hog1 directly activates enzymes involved in glycerol production (Westfall et al., 2008). Thirdly, Hog1 translocates into the nucleus and regulates, directly or indirectly, the expression of a large number of genes, including genes coding for glycerol producing enzymes (e.g. Gpd1; O'Rourke and Herskowitz, 2004). One can distinguish a short-term adaptation response, that rapidly promotes glycerol accumulation in the cell and a long-term adaptation response that results in preparing the cell to face prolonged periods of osmotic stress by triggering a large transcription program. Once the osmotic balance is restored, Hog1 is dephosphorylated and, in case of an excess in internal glycerol concentration, Fps1 channels reopen allowing glycerol to leak out (Beese et al., 2009). This pathway is schematically represented in Figure 1(left). For our application, we slightly modify the natural pathway. In addition to fusing Hog1 to a fluorescent marker that allows to detect its nuclear localization and hence quantify its activity, we fuse under the control of an osmo-responsive promoter the protein of interest, called X, to a second fluorescent marker that allows the quantification of the concentration of protein X.

Our motivation for using this pathway is threefold. Firstly, it has been extensively experimentally studied and quantitative models are available (Capaldi et al., 2008; Hersen et al., 2008; Klipp et al., 2005; Macia et al., 2009; Mettetal et al., 2008; Muzzey et al., 2009; Zi et al., 2010). Secondly, the output of the signal transduction pathway can be experimentally quantified. Indeed, if Hog1 is fused to a fluorescent protein, its nuclear localization can be quantified and provides a measure of the Hog1 activity: we have experimentally access to an important variable (Ferrigno et al., 1998). Thirdly, it has been experimentally shown that for fast osmolarity changes, the pathway integrates the signal: the transduction pathway acts as a low-pass filter with a bandwidth approximatively equal to 5×10^{-3} Hz (Hersen et al., 2008). This property allows us to emulate an analog control by rapidly switching (frequencies greater than 0.1 Hz) between two media: the normal growth medium and a sorbitol enriched ($\sim 1\,M$) medium. This is important since the microfluidic device we use (see below) does not allow for mixing solutions, whereas our control approach uses a continuous input variable for the external osmolarity.

2.2 The integrated platform

Because in comparison to open loop control, closed loop control approaches are generally less sensitive to model uncertainties and can compensate for external disturbances, feedback control seems unavoidable for our application. However, feedback control implies a real-time requirement: *in vivo* measurements, image analysis, computation of control strategies, and actuation on the cell environment must all be integrated and performed faster than the typical response time of the system.

The control platform is represented in Figure 1 (right). Time lapse fluorescent microscopy allows for monitoring the HOG pathway activity and gene expression at the single cell level. Using the micro-fluidic device developed by Hersen and colleagues, the cellular environment can be precisely controlled (2008). Not only this device allows a fast and well-controlled change of the cellular environment, but also, it guarantees that with the exception of its osmolarity the cell environment is otherwise held constant. In its current state, the platform integrates a microscope controller, a microfluidic pressure controller, and software for image analysis and PI control (Uhlendorf et al., 2011).

3. CONTROL APPROACH

We propose to use a model based control strategy. Therefore, we first present a model of the system. Then, we present and motivate the three main features of our control strategy: using a two-layer control, pulse modulation and a model predictive control approach. Lastly, we propose a control algorithm implementing the proposed approach.

3.1 Model developments

Many models of the yeast stress response have been developed (Hersen et al., 2008; Klipp et al., 2005; Mettetal et al., 2008; Muzzey et al., 2009; Zi et al., 2010). Because only a very limited number of variables can be measured simultaneously *in vivo* in single cells, and to simplify as much as possible parameter search, and even more importantly, state estimation problems, we have chosen to use models that are as simple as possible. For signal transduction and short-term cell adaptation, it was shown in Mettetal et al. (2008) and in Muzzey et al. (2009) that a number of essential features of the hyperosmotic stress response, notably its perfect adaptation and is frequency dependence, can be captured by simple 2- or 3-variable

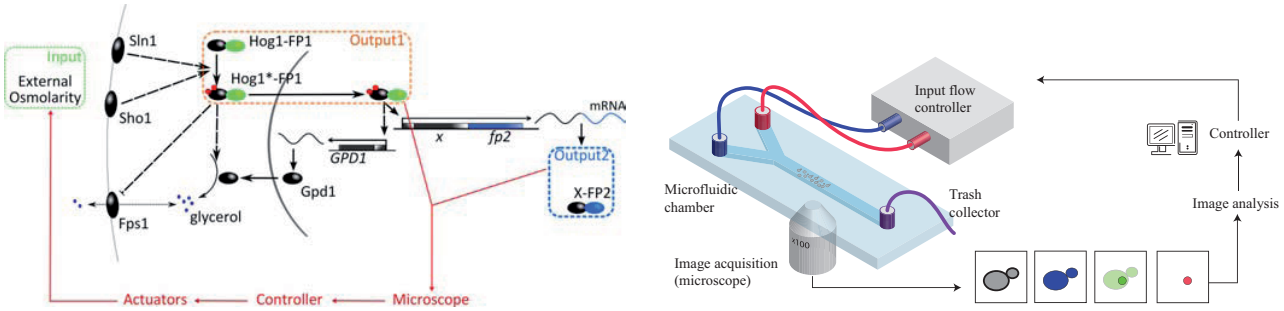


Fig. 1. **Left**: Schematic representation of the HOG pathway with natural and engineered feedbacks. Dashed arrows indicate indirect effects. For a detailed description, see the main text. FP1 and FP2 in the figure denote two different fluorescent proteins. **Right**: The integrated control platform. The main elements of the feedback loop are (i) a micro-fluidic device allowing a rapid control of the cellular environment, (ii) a microscope for phase contrast and fluorescence measurements, (iii) yeast cells with the protein of interest (X), Hog1, and a nuclear marker (Htb2) fused to compatible fluorescent proteins, and (iv) Matlab software for image analysis and controller implementation.

linear systems. Here, we propose the following switched linear model:

- if $osm_e \geq osm_i$: (hyperosmotic conditions)

$$\dot{osm}_i = \kappa_o hog - \gamma_o osm_i \quad (1)$$
$$\dot{hog} = \kappa_g(osm_e - osm_i) - \gamma_g hog \quad (2)$$

- if $osm_e < osm_i$: (hypoosmotic conditions)

$$\dot{osm}_i = \kappa_o hog - (\gamma_o + \gamma'_o) osm_i \quad (1')$$
$$\dot{hog} = -\gamma_g hog \quad (2')$$

In this model, osm_e, osm_i, and hog represent respectively the relative external osmolarity, that is the input, the relative internal osmolarity whose variations essentially result from glycerol synthesis and degradation/export, and the activity of the Hog1 protein, reflected by its nuclear localization. The relative osmolarity is defined as the difference between the normal osmolarity and the osmolarity at steady steate in the default conditions. Relative external and internal osmolarities therefore respectively correspond to the increase of osmolarity due to the addition of sorbitol in the growth media and due to the synthesis of glycerol in the cytoplasm. Like in the Mettetal and Muzzey models (2008; 2009), we assume that Hog1 activation is proportional to the intensity of the hyperosmotic stress, $osm_e - osm_i$. Using results from Mettetal et al. (2008), we neglect the Hog1-independant glycerol synthesis and simply assume that glycerol synthesis is proportional to Hog1 activity. We assume that the Fps1 glycerol channels are either open or closed, in hypoosmotic or hyperosmotic conditions, respectively. This yields the switched nature of the model. The term $\gamma'_o\, osm_i$ corresponds to glycerol diffusion out of the cell through Fps1 channels (assuming a negligible external glycerol concentration), whereas the term $\gamma_o\, osm_i$ corresponds to glycerol degradation in the cell. It was expected that $\gamma'_o \gg \gamma_o$. When fitting model parameters to experimental data generated in our lab, we indeed found that $\gamma'_o \gg \gamma_o \approx 0$.

For gene expression, we use the simple reaction-based model represented in Figure 2. In this model, mRNA synthesis is proportional to the nuclear concentration of the regulator of gene expression Hog1, and protein synthesis is proportional to mRNA concentration. mRNA and protein degradations are proportional to their concentrations (see

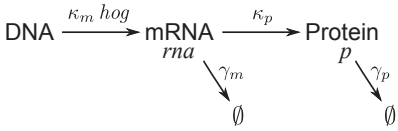


Fig. 2. A simple reaction-based model for gene expression.

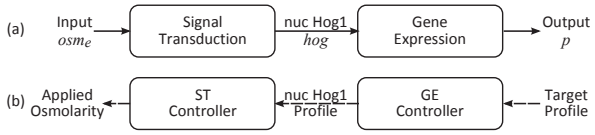


Fig. 3. **(a)** Decomposition of the osmotic stress response system into two subsystems: the relatively fast signal transduction system and the relatively slow gene expression system. Known and experimentally measurable quantities are represented. **(b)** Decomposition of the original control problem into two simpler control problems.

for example Wilkinson (2009) for a similar model). This reaction-based model can either be interpreted with a stochastic semantics using the Gillespie algorithm, or with a deterministic semantics, yielding the ODEs (3)-(4).

$$\dot{rna} = \kappa_m hog - \gamma_m rna \quad (3)$$
$$\dot{p} = \kappa_p rna - \gamma_p p \quad (4)$$

Here, p and rna represent respectively the concentrations of the target protein and of its messenger RNA. In this paper the stochastic and the deterministic semantics will be both used.

We will refer to the switched linear model (1)-(2)/(1')-(2') as the (fast) signal transduction model, and to the linear model (3)-(4) as the (slow) gene expression model. This gives rise to the schematic representation given in Figure 3(a), where we decompose the full system in two subsystems: one responsible for signal transduction, and one responsible for gene expression.

For the signal transduction model, we selected parameter values that minimize the mean square deviation between model prediction and experimental data produced in our lab (Uhlendorf et al., 2011), using a global optimization tool, implementing a covariance matrix adaptation evolution strategy (CMA-ES; Hansen and Ostermeier 2001). Parameter values are given in Table 1 (top).

κ_o	γ_o	γ'_o	κ_g	γ_g	osm_{i0}	hog_0
0.005	0	9.05	48	1.24	0	0.019

κ_m	γ_m	κ_p	γ_p	rna_0	p_0
0.1386	0.1386	0.3466	0.0347	0	0

Table 1. Parameter values and initial concentrations for the signal transduction (top) and gene expression (bottom) models.

For the gene expression model, and in absence of gene expression data, we used realistic parameter values. More precisely, we used degradation parameters values that correspond to mRNA and protein half-lives of 5 and 20 minutes (typically corresponding to a protein fused to a destabilized CFP (Hackett et al., 2006)). For production parameters, we used values that yield, in presence of a prolonged gene activation (i.e. $hog = 25$ in our context), a mean protein number of 250 and a standard deviation of 10% if one uses the stochastic interpretation of our gene expression reaction rate model (Ghaemmaghami et al. (2003); Friedman et al. (2006); see for example Wilkinson (2009) for the correspondence between the ODE and the discrete stochastic interpretations of reaction rate models). Note that the mean value for the number of protein X is somewhat arbitrary, since our target profiles are defined relatively to this value. Note also that protein folding time is not explicitly taken into account here. Parameter values are given in Table 1 (bottom).

3.2 Control strategy

The specific, 'cascaded' structure of our system and the possibility to observe the *hog* variable motivated us to decompose our original control problem into two simpler control problems (Figure 3b). Firstly, given the target protein profile, one looks for a desired Hog1 profile, considered as the input of the gene expression system. And secondly, given the desired Hog1 profile, considered this time as the output of the signal transduction system, one looks for external osmolarities to apply. In each case, only the model of the subsystem needs to be taken into account. This approach shares analogies with backstepping control strategies (Sepulchre et al., 1997).

Moreover, these two systems have different time scales. Indeed, response times for Hog1 signal transduction are typically in the order of one or 2 minutes, whereas response times for gene expression in yeast are much slower, typically on the order of 20 to 40 minutes. This allows us to develop a model predictive control strategy (Findeisen et al., 2007) in which the gene expression controller uses long-term predictions with large sampling times ($\approx$ 10 minutes), and the signal transduction controller uses short term predictions with high sampling times ($\approx$ 20 seconds).

A last important feature of the system is that cells adapt to osmotic stresses: natural negative feedback loops restore the cellular osmotic balance via glycerol production. For example, it has been experimentally demonstrated that to obtain a constant nuclear Hog1 concentration, and hence a constant protein synthesis rate, one has to apply a constantly increasing ('ramp') input (Muzzey et al., 2009). This fact, combined with the boundedness of the input implies that such strategies cannot be applied over long periods. Instead, we have experimentally shown - using a simple PI controller- that trapezoidal-like motifs of Hog1 nuclear localization can be repeatedly obtained, provided that these motifs are separated by sufficient time ($\approx$ 5 minutes in our conditions, Uhlendorf et al. (2011)). Therefore, we adopted a control strategy based on pulse-modulated signals for gene expression (Gelig and Churilov, 1998). This pulse-modulated approach exploits the fact that in absence of stress, glycerol channels open and let glycerol leak out of the cell, which effectively 'resets' the cell to its initial state. More precisely, Hog1 profiles will be defined by the number n of trapezoidal motifs, and each motif will be defined by the durations of its *off* state, d_{off}, and of its *on* state, d_{on}, as represented in Figure 4. To these times, we add fixed times for increase and decrease, and minimal durations of the *off* plateau, d_{relax}, and of the *on* plateau. For our computational studies, these fixed durations equal 2, 2, 5 and 1 minute, respectively. We also require that all d_{on} durations last less than 6 minutes to limit cell adaptation problems.

To summarize, we will develop a gene expression controller that, given a target protein profile, computes a desired pulse-modulated nuclear Hog1 profile (Figure 4), and a signal transduction controller that computes the osmolarity to apply to get the desired nuclear Hog1 profile (Figure 5).

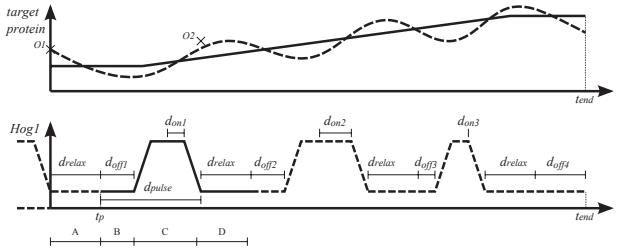


Fig. 4. Model predictive control for the GE controller. **(Bottom)** Hog1 profile with $n = 3$ motifs, defined by $d_{off} = (2, 2, 1, 4)$ and $d_{on} = (1, 2, 0)$ times. **(Top)** Corresponding protein profile (dotted line) with its reference profile (solid lines). This profile has been computed during period A, based on the observation $O1$. During periods B and D, no stress will applied ($osm_e = 0$), whereas in period C, the signal transduction controller will be used to apply osmotic stresses yielding the desired Hog1 profile. Based on measurement $O2$, the control strategy will be updated during period D. t_p denotes the start time of the current pulse.

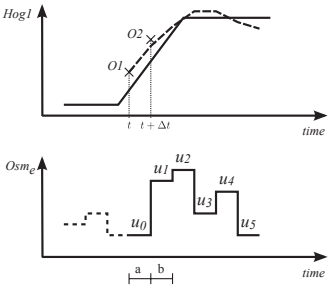


Fig. 5. Model predictive control for the ST controller. External osmolarity profile corresponding to $osm_e = (u_i)_{i\in[0,5]}$, computed during period a, based on observation $O1$. u_0 will be effectively applied during period b, and a new osmolarity profile will be computed based on observation $O2$.

3.3 Control algorithm

In this section, we propose a control algorithm implementing the control strategy presented above. The algorithm exploits two functions, searchHogProfile and searchOsmProfile, that we introduce first.

For the gene expression controller, a control strategy is defined by a nuclear Hog1 profile, that is, by the number of peaks n and the *on* and *off* durations of the peaks, $d_{on} \in \mathbb{R}^n_{\geq 0}$ and $d_{off} \in \mathbb{R}^{n+1}_{\geq 0}$, using the notations introduced in Figure 4. Then, given the target protein profile on a given time horizon (typically 100 minutes), the best control strategy at time t is the one that minimizes the mean square distance between the target protein profile and the protein profile that one obtains when applying the Hog1 profile on the input of subsystem 2, during the given time horizon. Good -although not necessarily optimal- solutions can be obtained by using global optimization tools that search for the best parameters n, d_{on} and d_{off}. This is implemented by the function $[n, d_{on}, d_{off}] = \text{searchHogProfile}(t, s_{ge}, target)$, with t the current time, s_{ge} the state of the gene expression subsystem at time t, and *target* the target protein profile. In practice, we consider a limited set of possible values for n, and for each value, we search for d_{on} and d_{off} parameters. We assume that these computations are executed in parallel, e.g. using a multicore processor.

For the signal transduction controller, a control strategy on the time horizon $[t, t + m\,\Delta t[$ is defined by m external osmolarity values $u = (u_1, \ldots, u_m) \in [0, 1]^m$. Given a desired nuclear Hog1 profile, the best control strategy is the one that minimizes the mean square error between the nuclear Hog1 profile obtained by applying the u_i osmolarities, each during Δt, and the desired nuclear Hog1 values, on the time interval $[t, t + m\,\Delta t[$. Again in practice, good solutions can be obtained by using global optimization tools that search for the best parameters $u_1, \ldots, u_m$. This is implemented by the function $u = \text{searchOsmProfile}(t, s_{st}, t_p, n, d_{on}, d_{off})$, with t the current time, s_{st} the state of the signal transduction subsystem at time t, and t_p, n, d_{on}, d_{off} defining the desired Hog1 profile, t_p being the start time of the first pulse.

Having defined the two functions searchHogProfile and searchOsmProfile, we can now describe our main algorithm (Algorithm 1). Note that the real-time requirement implies that measurements, image analysis, state estimation, and control strategy updates are terminated before the alloted time (d_{relax} for GE control and Δt for ST control).

4. TESTING THE CONTROL APPROACH ON SIMULATED DATA

Because testing control approaches on real cells is very time consuming, we first test the proposed approach on a simulated system. We present here these computational results.

We define the following two control problems that might be relevant for various applications.

- Problem 1: maintaining the concentration of protein X at a given level
- Problem 2: obtaining a sine wave temporal profile for the concentration of protein X

Solving the first problem *in vivo* would allow us to study the effect of well-controlled, steady perturbations, whereas solving the second control problem would allow us to study signal processing capabilities of gene networks, as it was done recently for signal transduction cascades (Hersen et al., 2008; Mettetal et al., 2008).

To test the robustness of our control approach with respect to the large variability of biological processes, we use a stochastic interpretation of the gene expression reaction rate model, and the Gillespie algorithm for the simulation of system behavior (Wilkinson, 2009). For the sake of simplicity, we ignore observation and state estimation problems. For real-life applications, these two issues also need to be appropriately solved.

In Figure 6, we present simulated results corresponding to control experiments for Problem 1 and 2, and using either a deterministic or a stochastic model for gene expression. For the stochastic cases, the examples provided are representative of typical results in terms of mean square deviation with respect to target profiles. More precisely, the root mean square (rms) distance for the solutions shown for Problem 1 and 2 are respectively 10.6 and 8.8, while the mean rms over 10 runs is respectively 11.6 and 8.5.

These results demonstrate the feasibility of the pulse modulated control strategy proposed, and show its robustness with respect to large biological variabilities, since for both problems, the concentration of the protein under control remains within admissible bounds around its target value. Indeed, even if these results seem unimpressive for traditional control problems, obtaining similar results *in vivo* would be a genuine tour de force.

Moreover, the proposed computation procedure, relying on global optimization to implement the MPC scheme, conforms with our real-time requirement, in the sense that, for each iteration, the computational times of the searchOsmProfile and searchHogProfile procedures is less than 300 and 20 seconds, respectively.

5. DISCUSSION

In this work, we have presented a specific control problem: controlling *in vivo* the expression of a protein such that its cellular concentration follows a target temporal profile. This control problem is directly motivated by current research in our lab. We have also proposed a control strategy tailored to the specificities of our biological application. It's main ingredients are the use of a two-layer model predictive control approach and of a pulse-modulated strategy for the control of gene expression. Using a simple switched linear model of the system, we have shown that the proposed approach satisfies our real-time requirements and is robust with respect to the significant noisiness of gene expression processes.

To the best of our knowledge, the specific problem of the precise control of the expression of a single gene has not been previously studied. However, several approaches have been developed for the control of gene networks.

Control Algorithm 1 Each iteration in the external *while* loop corresponds to applying one pulse and computing the next one, and each iteration in the internal *while* loop corresponds to applying an osmotic stress during a small Δt and computing the next one. Tasks performed in a synchronous manner are represented by a || sign. t denotes the current time.

// initialization prior to control experiment, using known parameters *target* and s^0_{ge}.
$[n, d_{on}, d_{off}] = \text{SEARCHHOGPROFILE}(0, s^0_{ge}, target)$
// control experiment starts
while $n \neq 0$ **do**
  $t_p = t$
  // B: no stress during $[t_p, t_p + d_{off1}[$
  $u_1 = 0$; **apply** u_1; **wait for** d_{off1};
  // C: active control during time interval $[t_p + d_{off1}, t_p + d_{pulse}]$
  while $t + \Delta t < t_p + d_{pulse}$ **do**
    apply u_1; **wait for** Δt || **measure** s_{st}; $[u_1, \ldots, u_m] = \text{SEARCHOSMPROFILE}(t, s_{st}, t_p, n, d_{on}, d_{off})$
  end while
  // D: no stress and global strategy update during time interval $[t_p + d_{pulse}, t_p + d_{pulse} + d_{relax}]$
  apply 0; **wait for** d_{relax} || **measure** s_{ge}; $[n, d_{on}, d_{off}] = \text{SEARCHHOGPROFILE}(t, s_{ge}, target)$
end while
apply $u = 0$; **wait for** d_{off1}
// control experiment ends with $t = t_{end}$

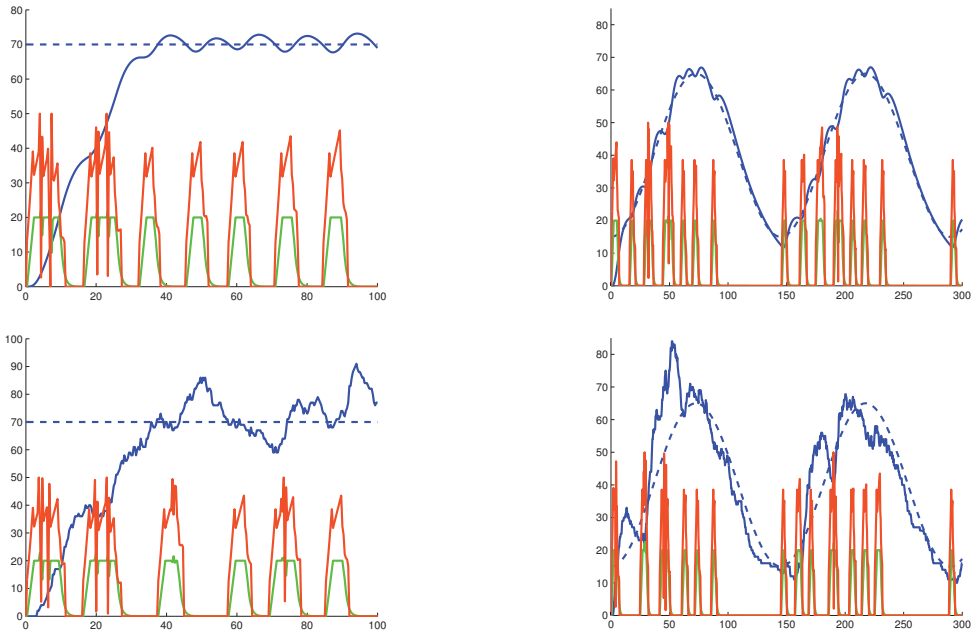


Fig. 6. Simulated control experiments. In all plots, dotted blue, orange, green, and solid blue represent the target protein profile, the applied osmolarity (x50), the nuclear concentration of Hog1, and the concentration of the X protein. System behavior is either computed using a deterministic (ODE, top) or a stochastic (Gillespie, bottom) model. Time is in minutes.

These approaches differ notably by the class of models that is considered. One can notably mention the works of Chaves and Gouzé (2011) and of Edwards et al. (2010), developed for qualitative piecewise affine models, of Datta et al. (2007), for probabilistic Boolean models, of Yu et al. (2010) for Sum models, and of Tumova et al. (2010), for piecewise affine models. Although we use here a simple switched linear model (actually being in the class studied by Tumova et al. (2010)), our MPC approach does not exploit any specific mathematical property of the model. In particular, we did not exploit the (piecewise) linear nature of the model. Because the predictive power of our model will be essential to get good performance for the control of the real biological system, being not bound to any specific mathematical formalism is likely to be a significant asset.

Further works naturally involve applying the proposed framework to the control of the actual biological system. Iterations between experiments, model developments and control strategy improvements will likely be needed.

ACKNOWLEDGEMENTS

We thank Pierre-Alexandre Bliman, Samuel Bottani, Madalena Chaves, Eugenio Cinquemani, and Francois Fages for helpful discussions and pointers to useful references.

REFERENCES

Beese, S.E., Negishi, T., and Levin, D. (2009). Identification of positive regulators of the yeast Fps1 glycerol channel. *PLoS Genetics*, 5(11), e1000738.

Capaldi, A., Kaplan, T., Liu, Y., Habib, N., Regev, A., Friedman, N., and O'Shea, E.K. (2008). Structure and function of a transcriptional network activated by the MAPK Hog1. *Nature Genetics*, 40(11), 1300--1306.

Chaves, M. and Gouzé, J.L. (2011). Exact control of genetic networks in a qualitative framework: the bistable switch example. *Automatica*. In press.

Colman-Lerner, A., Gordon, A., Pesce, C., Serra, E., Chin, T., Resnekov, O., Endy, D., and Brent, R. (2005). Regulated cell to cell variation in a cell fate decision system. *Nature*, 437(7059), 699--706.

Datta, A., Pal, R., Choudhary, A., and Dougherty, E. (2007). Control approaches for probabilistic gene regulatory networks. *IEEE Signal Processing Magazine*, 24(1), 54--63.

Edwards, R., Kim, S., and van den Driessche, P. (2010). Control design for sustained oscillation in a two gene regulatory network. *Journal of Mathematical Biology*, 62(4), 453--478.

Ferrigno, P., Posas, F., Koepp, D., Saito, H., and Silver, P. (1998). Regulated nucleo/cytoplasmic exchange of HOG1 MAPK requires the importin beta homologs NMD5 and XPO1. *EMBO Journal*, 17(19), 5606--5614.

Findeisen, R., Allögwer, F., and Biegler, L. (2007). *Assessment and Future Directions of Nonlinear Model Predictive Control*, volume 358 of *LNCIS*. Springer.

Friedman, N., L.Cai, and Xie, X. (2006). Linking stochastic dynamics to population distribution: an analytical framework of gene expression. *Physical Review Letters*, 97, 168302.

Gelig, A. and Churilov, A. (1998). *Stability and Oscillations of Nonlinear Pulse-Modulated Systems*. Birkhauser.

Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R., Belle, A., Dephoure, N., O'Shea, E., and Weissman, J. (2003). Global analysis of protein expression in yeast. *Nature*, 425(6959), 737--741.

Hackett, E., Esch, R., Maleri, S., and Errede, B. (2006). A family of destabilized cyan fluorescent proteins as transcriptional reporters in s. cerevisiae. *Yeast*, 23(5), 333--349.

Hansen, N. and Ostermeier, A. (2001). Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2), 159--195.

Hersen, P., McClean, M.N., Mahadevan, L., and Ramanathan, S. (2008). Signal processing by the HOG MAP kinase pathway. *Proceedings of the National Academy of Sciences of the USA*, 105(20), 7165--7170.

Hohmann, S. and Mager, W. (2003). *Yeast stress responses*. Topics in Current Genetics. Springer.

Klipp, E., Nordlander, B., Kruger, R., Gennemark, P., and Hohmann, S. (2005). Integrative model of the response of yeast to osmotic shock. *Nature Biotechnology*, 23(8), 975--982.

Macia, J., Regot, S., Peeters, T., Conde, N., Solé, R., and Posas, F. (2009). Dynamic signaling in the Hog1 MAPK pathway relies on high basal signal transduction. *Science Signaling*, 2(63), ra13.

Mettetal, J., Muzzey, D., Gomez-Uribe, C., and van Oudenaarden, A. (2008). The frequency dependence of osmo-adaptation in Saccharomyces cerevisiae. *Science*, 319(5862), 482--484.

Muzzey, D., Gomez-Uribe, C., Mettetal, J., and van Oudenaarden, A. (2009). A systems-level analysis of perfect adaptation in yeast osmoregulation. *Cell*, 138(1), 160--171.

Muzzey, D. and van Oudenaarden, A. (2009). Quantitative time-lapse fluorescence microscopy in single cells. *Annual Review of Cell and Developmental Biology*, 25(1), 301--327.

Newman, J., Ghaemmaghami, S., Ihmels, J., Breslow, D., Noble, M., DeRisi, J., and Weissman, J. (2006). Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature*, 441(7095), 840--846.

O'Rourke, S. and Herskowitz, I. (2004). Unique and redundant roles for HOG MAPK pathway components as revealed by whole-genome expression analysis. *Molecular Biology of the Cell*, 15(2), 532--542.

Sepulchre, R., Jankovic, M., and Kokotovic, P. (1997). *Constructive Nonlinear Control*. Springer-Verlag.

Tumova, J., Yordanov, B., Belta, C., Černá, I., and Barnat, J. (2010). A symbolic approach to controlling piecewise affine systems. In *IEEE Conference on Decision and Control (CDC'10)*, 4230--4235.

Uhlendorf, J., Bottani, S., Fages, F., Hersen, P., and Batt, G. (2011). Towards real-time control of gene expression: controlling the HOG signaling cascade. In *Pacific Symposium on Biocomputing, PSB'11*, volume 16, 338--349.

Westfall, P., Patterson, J., Chen, R., and Thorner, J. (2008). Stress resistance and signal fidelity independent of nuclear MAPK function. *Proceedings of the National Academy of Sciences of the USA*, 105(34), 12212--12217.

Wilkinson, D. (2009). Stochastic modelling for quantitative description of heterogeneous biological systems. *Nature Reviews Genetics*, 10(2), 122--133.

Yu, W., Lu, J., Wang, Z., Cao, J., and Zhou, Q. (2010). Robust H infinite control and uniformly bounded control for genetic regulatory network with stochastic disturbance. *IET Control Theory & Applications*, 4(9), 1687--1706.

Zi, Z., Liebermeister, W., and Klipp, E. (2010). A quantitative study of the Hog1 MAPK response to fluctuating osmotic stress in S. cerevisiae. *PLoS ONE*, 5(3), e9522.

B.3 Appended paper 3: Long-term model predictive control of gene expression at the population and single-cell levels

Long-term model predictive control of gene expression at the population and single-cell levels

Jannis Uhlendorf, Agnès Miermont, Thierry Delaveau, Gilles Charvin, François Fages, Samuel Bottani, Gregory Batt and Pascal Hersen

Proceedings of the National Academy of Sciences, 109(35):14271-14276, 2012

Long-term model predictive control of gene expression at the population and single-cell levels

Jannis Uhlendorf[a,b], Agnès Miermont[b], Thierry Delaveau[c], Gilles Charvin[d], François Fages[a], Samuel Bottani[b], Gregory Batt[a,1,2], and Pascal Hersen[b,e,1,2]

[a]Contraintes Research Group, Institut National de Recherche en Informatique et en Automatique, INRIA Paris-Rocquencourt, 78150 Rocquencourt, France; [b]Laboratoire Matière et Systèmes Complexes, Unité Mixte de Recherche 7057 Centre National de la Recherche Scientifique and Université Paris Diderot, 75013 Paris, France; [c]Laboratoire de Génomique des Microorganismes, Unité Mixte de Recherche 7238 Centre National de la Recherche Scientifique and Université Pierre et Marie Curie, 75006 Paris, France; [d]Institut de Génétique et Biologie Moléculaire et Cellulaire, 67400 Illkirch, France; and [e]Mechanobiology Institute, National University of Singapore, Singapore 117411

Edited by David A. Weitz, Harvard University, Cambridge, MA, and approved July 16, 2012 (received for review April 23, 2012)

Gene expression plays a central role in the orchestration of cellular processes. The use of inducible promoters to change the expression level of a gene from its physiological level has significantly contributed to the understanding of the functioning of regulatory networks. However, from a quantitative point of view, their use is limited to short-term, population-scale studies to average out cell-to-cell variability and gene expression noise and limit the nonpredictable effects of internal feedback loops that may antagonize the inducer action. Here, we show that, by implementing an external feedback loop, one can tightly control the expression of a gene over many cell generations with quantitative accuracy. To reach this goal, we developed a platform for real-time, closed-loop control of gene expression in yeast that integrates microscopy for monitoring gene expression at the cell level, microfluidics to manipulate the cells' environment, and original software for automated imaging, quantification, and model predictive control. By using an endogenous osmostress responsive promoter and playing with the osmolarity of the cells environment, we show that long-term control can, indeed, be achieved for both time-constant and time-varying target profiles at the population and even the single-cell levels. Importantly, we provide evidence that real-time control can dynamically limit the effects of gene expression stochasticity. We anticipate that our method will be useful to quantitatively probe the dynamic properties of cellular processes and drive complex, synthetically engineered networks.

model based control | computational biology | high osmolarity glycerol pathway | quantitative systems biology

Understanding the information processing abilities of biological systems is a central problem for systems and synthetic biology (1–6). The properties of a living system are often inferred from the observation of its response to static perturbations. Time-varying perturbations have the potential to be much more informative regarding the dynamics of cellular functions (7–12). Currently, it is not possible to precisely perturb protein levels in an analogous manner, even though this perturbation would be instrumental in our understanding of gene regulatory networks. Indeed, despite the development of novel regulatory systems, including various RNA-based solutions (13), transcriptional control by means of inducible promoters is still the preferred method for manipulating protein levels (14, 15). Unfortunately, inducible promoters have several generic limitations. First, there is a significant delay between gene expression activation and effective protein synthesis. Second, many cellular processes can interfere with gene expression through internal feedback loops whose effects are hard to predict. Third, the process of gene expression shows significant levels of noise (16–18). Given these limitations, novel experimental strategies are required to gain quantitative, real-time control of gene expression in vivo.

Here, we see the problem of manipulating gene expression to obtain given temporal profiles of protein levels as a model-based control problem. More precisely, we investigate the effectiveness of computerized closed-loop control strategies to control gene expression in vivo. In model-based closed-loop control, a model of the system is used to constantly update the control strategy based on real-time observations. We propose an experimental platform that implements such an in silico closed loop in the budding yeast *Saccharomyces cerevisiae*. We show that gene expression can be controlled by repeatedly stimulating a native endogenous promoter over many cell generations (>15 h) for both time-constant and time-varying target profiles and at both the population and single-cell levels. Recently, Milias-Argeitis et al. (19) also proposed an approach for feedback control of gene expression in yeast. In contrast to their work, we propose a method that is effective at the single-cell level, for time-varying target profiles, and robust despite the presence of strong internal feedback loops. We start by describing the gene induction system and the experimental platform before discussing its efficiency.

Results and Discussion

Controlled System. We based our approach on the well-known response of yeast to an osmotic shock, which is mediated by the high osmolarity glycerol (HOG) signaling cascade. Its activation leads to the phosphorylation of the protein Hog1 (Fig. 1*A*), which orchestrates cell adaptation through glycerol accumulation. Phosphorylated Hog1 promotes glycerol production by activating gene expression in the nucleus as well as stimulating glycerol-producing enzymes in the cytoplasm. After they are adapted, the cells do not sense the hyperosmotic environment anymore, the HOG cascade is turned off, and the transcriptional response stops (20–22). In control terms, yeast cells implement several short-term (non transcriptional) and long-term (transcriptional) negative feedback loops that ensure perfect adaptation to the osmotic stress (10, 23). Because of these adaptation mechanisms, it is a priori challenging to control gene expression induced by osmotic stress. It is, thus, an excellent system to show that one can robustly control protein levels, even in the presence of internal negative feedback loops. Several genes are up-regulated in response to a hyperosmotic stress. These genes include the nonessential gene *STL1*, which codes for a glycerol proton symporter (24, 25). We decided to use its native promoter (pSTL1) to drive the expression of yECitrine, a fluorescent reporter. Applying an osmotic stress transiently activated the HOG cascade (Fig. 1*B*), and yECitrine levels reached modest values (600 fluorescence units) (Fig. 1*B*). Importantly, when short but repeated stresses were applied, pSTL1 could be repeatedly activated, and much higher levels could be reached (Fig. 1*C*).

Author contributions: J.U., G.B., and P.H. designed research; J.U. performed research; J.U., A.M., T.D., G.C., F.F., S.B., G.B., and P.H. contributed new reagents/analytic tools/software; J.U., G.B., and P.H. analyzed data; and J.U., G.B., and P.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence may be addressed. E-mail: gregory.batt@inria.fr or pascal.hersen@univ-paris-diderot.fr.

[2]G.B. and P.H. contributed equally to this work.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1206810109/-/DCSupplemental.

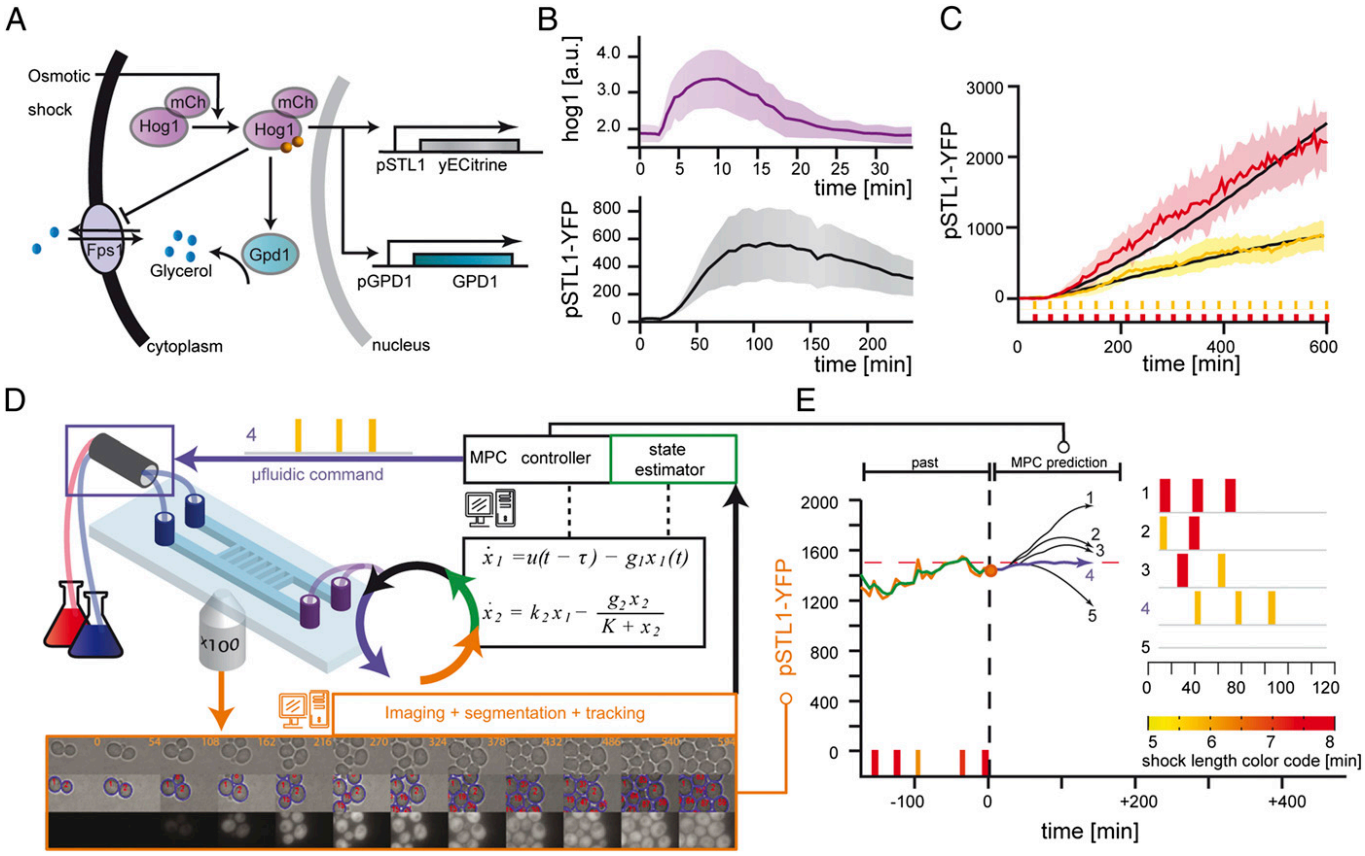


Fig. 1. A platform for real-time control of gene expression in yeast. (*A*) A hyperosmotic stress triggers the activation and nuclear translocation of Hog1. Short-term adaptation is mainly implemented by cytoplasmic activation of the glycerol-producing enzyme Gpd1 and closure of the aqua-glyceroporin channel Fps1. Long-term adaptation occurs primarily through the production of Gpd1. (*B*) When maintained in a hyperosmotic environment (1 M sorbitol), the HOG cascade was quickly activated, which is seen by Hog1 nuclear enrichment. This transient signaling response lasted typically <20 min. The expression level of pSTL1-yECitrine (YFP) increased after an ~20-min delay, peaked around 600 fluorescence units after 100 min, and then decayed. (*C*) In contrast, the fluorescence level showed a continuous increase when stimulated periodically (T = 30 min). The increase rate was larger for longer pulses (red, 8 min; yellow, 5 min). Black curves are the expected behaviors based on our model of the pSTL1 induction. Solid lines and their envelopes are the experimental means and SDs of the cells' fluorescence. (*D*) Yeast cells grew as a monolayer in a microfluidic device that was used to rapidly change the cells' osmotic environment (blue frame) and image their response. Segmentation and cell tracking were done using a Hough transform (orange frame). The measured yECitrine fluorescence, either of a single cell or of the mean of all cells, was then sent to a state estimator connected to an MPC controller. A model (black frame) of pSTL1 induction was used to find the best possible series of osmotic pulses to apply in the future so that the predicted yECitrine level follows a target profile. (*E*) At the present time point (orange circle), the system state is estimated (green), and the MPC searches for the best input (pulse duration and number of pulses) (see text and *SI Materials and Methods*), which minimizes the distance of the MPC predictions (black curves) to the target profile (red dashed line) for the next 2 h. Here, the osmotic series of pulses that corresponds to the blue curve (4) was selected and sent to the microfluidic command. This control loop is iterated every 6 min.

A closed-loop control of the pSTL1 activity requires the acquisition and analysis of live cell images, the computation of the input (i.e., osmolarity) to be applied in the near future, and the ability to change the cells osmotic environment accordingly (Fig. 1 *D* and *E*).

Experimental Platform. To observe the cells and control their environment, we designed a versatile platform made of standard microscopy and microfluidic parts. The microfluidic device contained several 3.1-μm-high chambers that were connected by both ends to large channels through which liquid media could be perfused (Fig. 1*D*). Because the typical diameter of an *S. cerevisiae* cell is 4–5 μm, the cells were trapped in the chamber and grew as a monolayer. Their motion was limited to slow lateral displacement due to cell growth (Fig. S1). This design allowed for long-term cell tracking (>15 h) and relatively rapid media exchanges (~2 min). The HOG pathway was activated by switching between normal and sorbitol-enriched (1 M) media.

Model of pSTL1 Induction. To decide what osmotic stress to apply at a given time, we used an elementary model of pSTL1 induction. Many models have been proposed for the hyperosmotic stress response in yeast (10, 26–30). We used a generic model of gene expression written as a two-variable delay differential equation system, where the first variable denotes the recent osmotic stress felt by the cell and the second variable is the protein fluorescence level (Fig. 1*D*, *Materials and Methods*, Table S1, and *SI Materials and Methods*). Because our goal was to show robust control, despite the presence of unmodeled feedback loops, the adaptation mechanisms described above were purposefully neglected. The choice of this model was also motivated by the tradeoff between its ability to quantitatively predict the system's behavior (favors complexity) and the ease of solving state estimation problems (favors simplicity). Despite its simplicity, we found a fair agreement between model predictions and calibration data corresponding to fluorescence profiles obtained by applying either isolated or repeated osmotic shocks of various durations (Fig. 1*C* and Fig. S2).

Closing the Loop. The fluorescence intensity of a single cell arbitrarily chosen at the start of the experiment, or the average fluorescence intensity of the cell population, was sent to a state estimator (extended Kalman filter discussed in *SI Materials and Methods*) connected to a model predictive controller (31). Model Predictive Control (MPC) is an efficient framework well-adapted

to constrained control problems. Schematically, given a model of the system and desired temporal profiles for the system's outputs, MPC aims at finding inputs to minimize the deviation between the outputs of the model and the desired outputs. The control strategy is applied for a (short) period. Then, the new state of the system is observed, and this information is used to compute the control strategy to be applied during the next time interval. This receding horizon strategy yields an effective feedback control. In practice, every 6 min, given the current estimate of the system state, past osmotic shocks, and our model of gene expression, the controller searched for the optimal number of osmotic pulses to apply within the next 2 h and their optimal start times and durations (Fig. 1*E*). If a shock had to be applied within the next 6 min, then it was applied. Otherwise, the same computation was reiterated 6 min later based on new observations, thus effectively closing the feedback loop. Here, we dealt with short-term cell adaptation by imposing a maximal stress duration of 8 min and a 20-min relaxation period between consecutive shocks. Under such conditions, cells stay responsive to osmotic stress at all times (Fig. 1*C*). This can be explained by the fact that, in absence of stress, the glycerol channel Fps1 opens (21, 32) and lets the glycerol leak out of the cell, thus effectively resetting the osmotic state of the cells (29).

Note that a proportional integral (PI) controller would have been an attractive alternative, because it would not have required the development of a model of the system. With PI controllers, the applied input (i.e., stress) is simply the weighted sum of the current error (deviation between target and measured outputs) and the integral of the (recent) past errors. Consequently, using a PI controller to reach high levels of fluorescence would lead to a control strategy in which high stress is maintained over extended periods of time. This condition would trigger cell adaptation and eventually lead to a stalled situation in which the maximal stress is applied without any effect.

Closed-Loop Population Control Experiments. Our first goals were to maintain the average fluorescence level of a cell population at a given constant value (set-point experiment) and force it to follow a time-varying profile (tracking experiment). Both types of experiments lasted at least 15 h, starting with a few cells and ending with 100–300 cells in the field of view (Fig. S3). The control objective was to minimize the mean square deviations (MSDs) between the mean fluorescence of the population of cells and the target profile. We succeeded in maintaining the average fluorescence level at a given constant value or forcing it to follow several given time-varying profiles (Fig. 2 *A–D*, Figs. S3, S4, and S5, and Movies S1, S2, and S3). Admissible time-varying target profiles were obviously constrained by the intrinsic timescales of the system, such as the maximal protein production and degradation rates. However, within these constraints, graded responses could be obtained. In Fig. 2*C*, for example, the trapeze slope is less steep than what maximal pSTL1 induction can deliver (Fig. 2 *A* and *B*). Note that our control strategy opened the possibility to reach higher fluorescence levels than what full induction with a step shock would allow (compare with Fig. 1*B*). Indeed, because of cell perfect adaptation to hyperosmotic stresses, a sustained 1 M sorbitol shock triggers only a transient gene expression and fluorescence peaks at moderate levels (Fig. 1*B*). By using repeated, well-separated pulses,

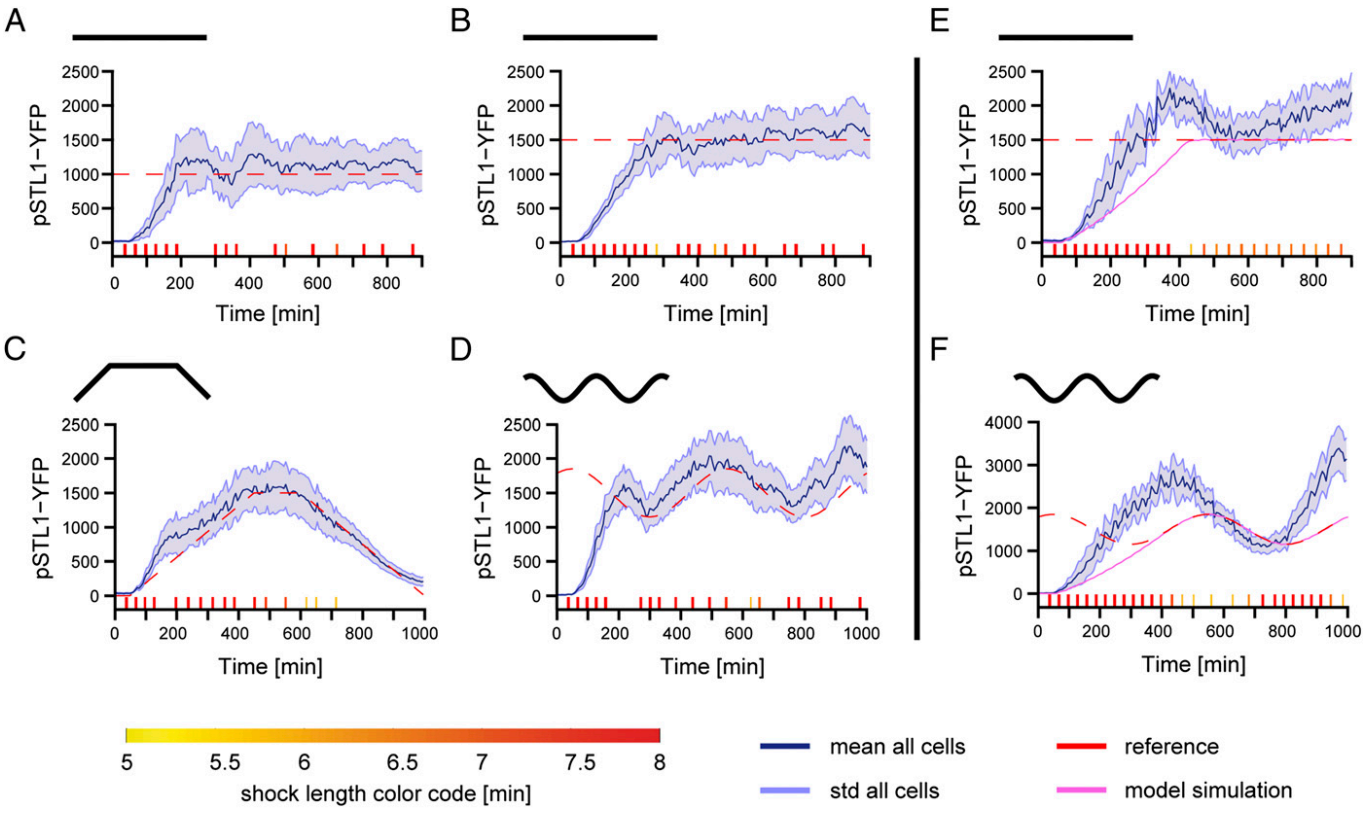


Fig. 2. Real-time control of gene expression can be achieved at the population level. (*A* and *B*) Set-point control experiments with target values 1,000 and 1,500 fluorescence units (f.u.; red dashed line). This unit is the same across all graphs (no renormalization). To avoid desensitizing the HOG pathway, the controller repeatedly applied short osmotic pulses (durations between 5 and 8 min). The timeline of osmotic events is shown at the bottom of each graph (color code along the bottom). Shock starting times and durations were computed in real time. The measured mean cell fluorescence is shown as solid blue lines. The envelopes indicate SD of the fluorescence distribution across the yeast population. (*C* and *D*) Tracking control experiments. In *C*, the target has a trapezoidal shape (maximum at 1,500 f.u.). In *D*, the target is sinusoidal (average value at 1,500 f.u.). In both cases, the mean level of fluorescence successfully follows the time-varying target profile. (*E* and *F*) Open-loop control experiments. Two examples of open-loop control (the osmotic inputs were computed using our model before starting the experiments) showing poor control quality. Errors accumulate over time. The simulated behavior of the system is represented in violet.

pSTL1 was iteratively activated (Fig. 1*C* and Movie S4). To assess the effective control range, we performed additional control experiments with target values spanning an order of magnitude (200–2,000 fluorescence units) (Fig. S5). Despite an initial overshoot for the lower target (200 fluorescence units), our results showed good control accuracy over time.

Quantitative limitations of our experimental platform can originate from the model, the state estimator, the control algorithm, and the intrinsic biological variability of gene expression. In silico analysis showed that applying the proposed control strategy to the (estimated state of the) system resulted in control performances that were significantly better than those obtained experimentally (Fig. 2 *E* and *F* and Fig. S4). Therefore, the control algorithm performed well, and future improvements should focus on system modeling and state estimation to better represent the experimental state of the system. To assess the importance of biological variability and modeling limitations, we carried out open-loop control experiments with the same objectives and the same model of the system. A time series of osmotic pulses was computed before the experiment and then sent to yeast cells without performing real-time corrections. Important deviations were found, indicating clear discrepancies between model predictions and the long-term system behavior (Fig. 2 *E* and *F*). As expected, open-loop strategies cannot result in a quantitative, robust control of gene expression. In contrast, closed-loop control performs well, despite significant biological variability and/or limited model accuracy.

Closed-Loop Single-Cell Control Experiments. In a second set of experiments, we focused on the real-time control of gene expression at the single-cell level. We tracked one single cell over at least 15 h and used its fluorescence to feed the MPC controller. As shown in Fig. 3, we obtained results with quality that is out of reach of any conventional gene induction system, both for constant and time-varying target profiles (Movies S5, S6, and S7). Because of intrinsic noise in gene expression, single-cell control was a priori more challenging than population control. Indeed, compared with the mean fluorescence levels in population control experiments, the fluorescence levels of controlled cells in single-cell control experiments showed larger fluctuations around the target values. However, at the cell level, the MSDs of controlled cells obtained in single-cell control experiments were significantly smaller than the MSDs of a cell in population control experiments (Fig. 4*B*, *SI Materials and Methods*, Table S2, and Fig. S6). For set-point control experiments in which fluctuations happen around a fixed reference value, we also defined the fluorescence noise level as the standard deviation (SD) over the mean. Again, we found that single-cell control significantly decreased noise at the cell level (Fig. 4*C*, *SI Materials and Methods*, Table S2, and Fig. S6). Taken together, these results show that real-time control effectively improves control quality and counteracts the effects of noise in gene expression when performed at the single-cell level. Interestingly, single-cell control experiments showed that, in few cases, the controlled cell behaved significantly differently from the rest of the population over extended periods of time (e.g., see Fig. 3*A*), suggesting long-term memory effects for gene expression spanning many cell generations. Lastly, the fact that, for different controlled cells but the same control objective, the decisions of the closed-loop controller were markedly different (Fig. 3*E*) highlights the fact that feedback control was critical to achieve good control performance at the single-cell level. This suggests that cell-to-cell variability and noise in gene expression fundamentally limit the quality of any open-loop inducible system.

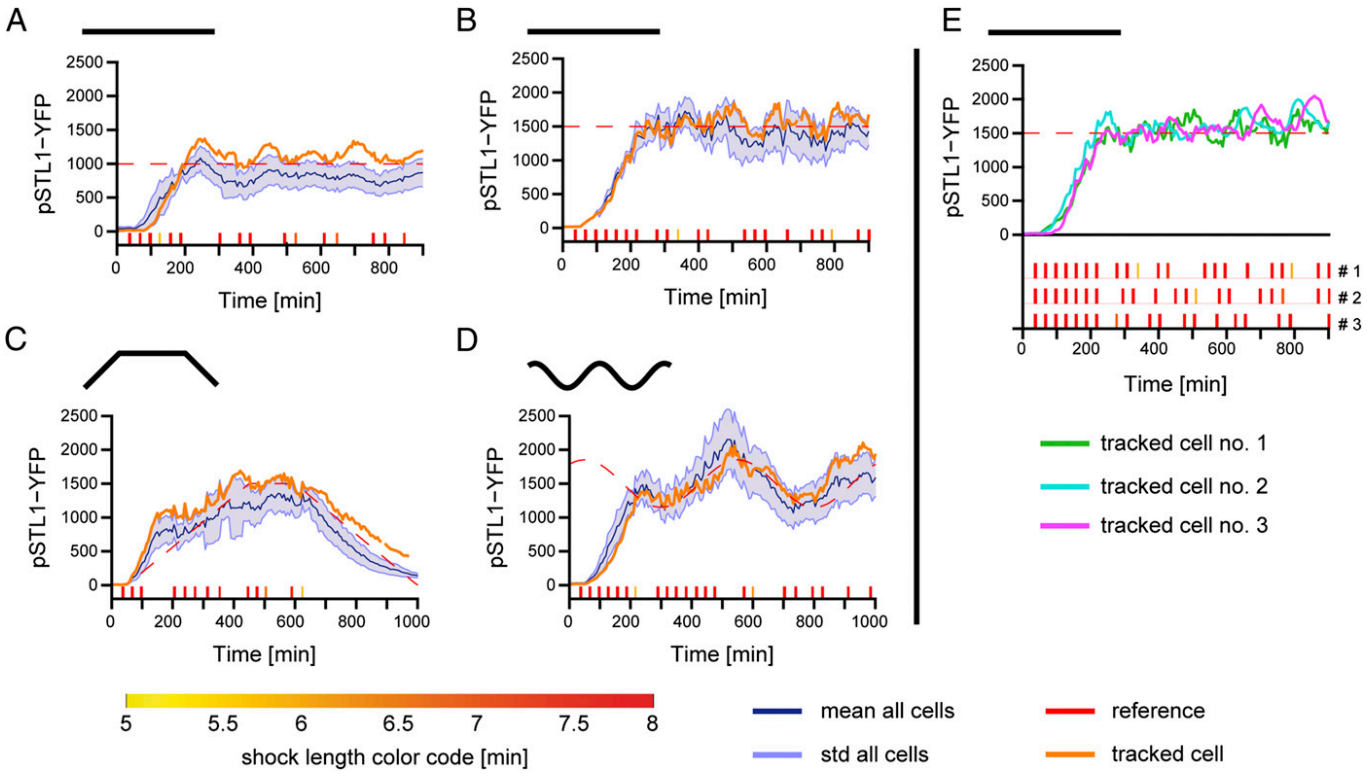


Fig. 3. Real-time control of gene expression can be achieved at the single-cell level. (*A* and *B*) Set-point control experiments at values 1,000 and 1,500 f.u. The yECitrine fluorescence of the controlled cells is shown as orange lines. The blue line and its envelope indicate the mean fluorescence and the SD of the fluorescence across the cell population. The population follows the target profile but with less accuracy than the controlled single cell. (*C* and *D*) Tracking control experiments. In *C*, the target has a trapezoidal shape (maximum at 1,500 f.u.). In *D*, the target is sinusoidal (average at 1,500 f.u.). (*E*) The fluorescence of the controlled cell in three different single-cell control experiments is represented together with the osmolarity profiles that were applied. Different experiments are labeled with different colors, and therefore, their corresponding osmotic inputs can be identified. It appears that, for each cell, the controller decisions were markedly different, showing that cell-to-cell variability was at play and that feedback control was critical when performing single-cell control.

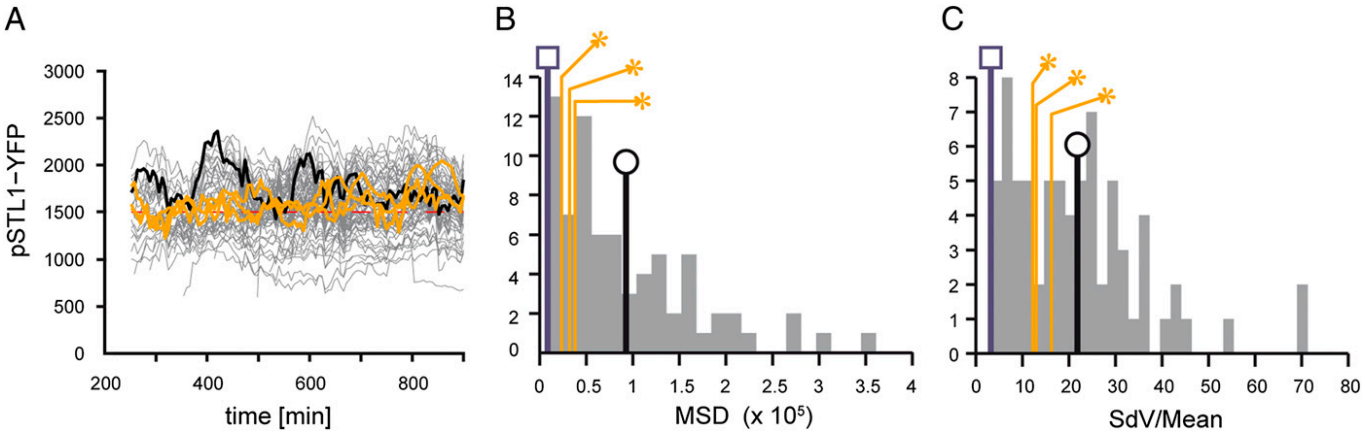


Fig. 4. Effectiveness of closed-loop control. (*A*) Single-cell fluorescence time profiles in two population control experiments (thin gray lines) and three single-cell control experiments (thick orange lines). One representative trace of a single cell in a population control experiment is shown in black. (*B*) Distribution of the MSDs of individual cells in population control experiments (gray). MSDs are defined with respect to the target profiles. The orange bars (stars) show the MSDs for the controlled cells in three single-cell control experiments. These data are compared with the mean MSD of single cells when controlling the population (black line, circle), which shows lower control quality. As expected, the control quality of the population is better (blue line, square), because noise in gene expression is averaged out. (*C*) Distribution of the noise levels defined as the ratio of the SD to the mean. Lower noise levels are observed for controlled cells in single-cell control experiments (orange, star) than a random cell in population control experiments (black line).

Conclusions

We showed that gene expression can be controlled in real time with quantitative accuracy at both the population and single-cell levels by interconnecting conventional microscopy, microfluidics, and computational tools. Importantly, we provided evidence that real-time control can dynamically limit the effects of gene expression stochasticity when applied at the single-cell level. This model predictive control framework overcame the presence of a significant delay between the environmental change and the fluorescent protein observation and the action of strong non-modeled endogenous negative feedback loops. The fact that good control results can be obtained in a closed-loop setting with a relatively coarse model of an endogenous promoter (compare with open-loop results) suggests that extensive modeling will not be required to transpose our approach to other endo- and exogenous induction systems (e.g., the galactose, methionine, or tetracycline inducible promoters). To appreciate the difficulty of the control problems that we addressed, one should keep in mind that the controlled system, a yeast cell, is an extremely complex and partially known dynamical system and that the controlled process, gene expression, is intrinsically stochastic.

Despite the fact that the importance of control theory for systems and synthetic biology has been widely recognized for more than a decade (33, 34), the actual use of in silico feedback loops to control intracellular processes has only been proposed recently. In 2011, we showed that the signaling activity in live yeast cells can be controlled by an in silico feedback loop (35). Using a PI controller, we controlled the output of a signal transduction pathway by modulating the osmotic environment of cells in real time. A similar framework has been proposed by Menolascina et al. (36). More recently, Toettcher et al. (37) used elaborate microscopy techniques and optogenetics to control (in real time and the single-cell level) the localization and activity of a signal transduction protein (PI3K) in eukaryotic cells. Interestingly, they were able to buffer external stimuli and clamp phosphatidylinositol (3,4,5)-triphosphate (PIP3) levels for short time scales. Because these two frameworks necessitate image acquisition at a high frequency, they are not suitable for long-term experiments. The most closely related work is the work by Milias-Argeitis et al. (19). Using optogenetic techniques, Milias-Argeitis et al. (19) managed to control the expression of a yeast gene to a constant target value over a few hours. Their approach is based on a chemostat culture and well-adapted for biotechnological applications, such as the production of biofuels or small-molecules. However, because it does not allow for single-cell tracking and control, it is less adapted to probe biological processes in single-cell quantitative biology applications. Controlling small cell populations, or even single cells, may be needed in multicellular systems, where cells differ by their genotype (38) or physical location (39).

Connecting living cells to computers is a promising field of research both for applied and fundamental research. By maintaining a system around specific operating points or driving it out of its standard operating regions, our approach offers unprecedented opportunities to investigate how gene networks process dynamical information at the cell level. We also anticipate that our platform will be used to complement and help the development of synthetic biology through the creation of hybrid systems resulting from the interconnection of in vivo and in silico computing devices.

Material and Methods

Yeast Strains and Plasmids. All experiments were performed using a pSTL1::yECitrine-HIS5, Hog1::mCherry-hph yeast strain derived from the S288C background. Cells were cultured overnight in synthetic complete (SC) medium at 30 °C; 4 h before loading them into the microfluidic chip, 60 μL overnight culture were diluted into 5 mL SC, thus obtaining an OD of ~0.19. During the experiment, cell growth continued, with a doubling time between 100 and 250 min (*SI Materials and Methods* and Movies S1, S2, S3, S4, S5, S6, and S7).

Microfluidics. We microfabricated a master wafer by standard soft lithography techniques. A microfluidic chip was made by casting polydimethylsiloxane (PDMS) (Sylgard 184 kit; Dow Corning) on the master wafer, curing it at 65 °C overnight, pealing it off, and bonding it to a glass coverslip after plasma activation. Cells were loaded into the imaging chamber by syringe injection. This created a positive pressure, which let the cells enter the trap. Liquid medium was flown by aspiration into the device using a peristaltic pump (IPC-N; Ismatec) placed after the microfluidic device. We used a flow rate of 230 μL/min. A computer-controlled three-way valve (LFA series; The Lee Company) was used to select between regular medium (SC) or the same medium supplemented with 1 M sorbitol. A switch of the valve state did not lead to an instantaneous change of the cells' environment inside the microfluidic device: a certain time (depending on the flow rate) was needed for the fluid to pass from the valve to the channels and the imaging chamber (Fig. S1).

Microscopy and Experimental Setup. We used an automated inverted microscope (IX81; Olympus) equipped with an X-Cite 120PC fluorescent illumination system (EXFO) and a QuantEM 512 SC camera (Roper Scientific). The YFP filters used were HQ500/20× (excitation filter; Chroma), Q515LP (dichroic; Chroma), and HQ535/30M (emission; Chroma). All these components were driven by the open-source software μManager (40), a plug-in of ImageJ (41), which we interfaced with Matlab using in-house–developed code. The temperature of the microscope chamber, which also contained the media reservoirs, was constantly held at a temperature of 30 °C by a temperature control

system (Life Imaging Services). Images were taken with a 100× objective (PlanApo 1.4 NA; Olympus). The fluorescence exposure time was 200 ms, with fluorescence intensity set to 50% of maximal power. The fluorescence exposure time was chosen such that the fluorescent illumination did not cause noticeable effects on cellular growth over extended periods of time. Importantly, illumination, exposure time, and camera gain were not changed between experiments, and no data renormalization was done. Therefore, the fluorescence intensities can be directly compared across experiments.

Image Analysis. The cellular boundaries were identified on the bright-field image using a circular Hough transform implemented in Matlab (42). For tracking, we compared the current image with the previous one, defined a cell-to-cell distance matrix, and used linear optimization to match pairs of cells. The tracking process was made more robust by also considering the last but one image if a gap was detected (caused by rare segmentation errors). The YFP fluorescence level in each cell was defined as the mean fluorescence level taken over the cell area after subtraction of the background fluorescent level. The signaling activity of the Hog1 cascade can be estimated by measuring the Hog1 nuclear enrichment. We defined the nuclear enrichment of Hog1::mCherry as the difference between the minimal and maximal fluorescence intensities within a cell. Maximal and minimal Hog1::mCherry intensities were computed by averaging the fluorescence of the 15 brightest and 15 dimmest pixels, respectively.

Modeling. The controller used a two dimensional ordinary differential equation (ODE) model to predict the behavior of the system:

$$\dot{x}_1 = u(t-\tau) - g_1 x_1$$

and

$$\dot{x}_2 = k_2 x_1 - g_2 \frac{x_2}{K + x_2},$$

where x_1 denotes the recent osmotic stress and x_2 denotes the protein fluorescence level. The osmotic input (u) is shifted by $\tau = 20$ min to account for the observed delay in the system. The remaining parameters have been estimated based on several calibration experiments: $g_1 = 4.02 \times 10^{-3}$, $k_2 = 0.58$, $g_2 = 37.5$, $K = 750$, and $\tau = 20$ (*SI Materials and Methods*, Table S1, and Fig. S2).

State Estimation. We implemented an extended Kalman filter, which estimates the system state based on fluorescent observations and the model of the system. The parameters of the filter (measurement noise R and process noise Q) were set to R = 2,500 and Q = diag(0.37, 925).

Model Predictive Control. The controller searches for osmolarity profiles that minimize the squared deviations between model output and target profile within the next 120 min, while fulfilling the input constraints (pulse duration of 5 to 8 min separated by at least 20 min). In practice, this problem is recast into a parameter search problem, in which parameters are used for encoding stress starting times and shock durations and solved using the global optimization tool CMAES. Because image analysis and parameter search may take up to 3 min, the input to be applied is not immediately available at the time of the measurement. Consequently, we apply at time t the input that was computed at time $t - 3$ min.

ACKNOWLEDGMENTS. The authors acknowledge discussions with D. di Bernardo (Tigem), P.-Y. Bourguignon (Max Planck), S. Léon (Institut Jacques Monod & Centre National de la Recherche Scientifique), F. Devaux, M. Garcia (Université Pierre et Marie Curie), J. M. di Meglio, A. Prastowo (Matière et Systèmes Complexes), R. Bourdais (Supélec), A. Kabla (Cambridge University), E. Cinquemani, H. de Jong, D. Ropers, J. Schaul, and S. Stoma (Institut National de Recherche en Informatique et en Automatique). We acknowledge the support of the Agence Nationale de la Recherche (under the references DiSiP-ANR-07-JCJC-0001 and ICEBERG-ANR-10-BINF-06-01), of the Région Ile de France (C'Nano-ModEnv), of the Action d'Envergure ColAge from INRIA/INSERM (Institut Nationale de la Santé et de la Recherche Médicale), of the MechanoBiology Institute, and of the Laboratoire International Associé CAFS (Cell Adhesion France-Singapour).

1. Bhalla US, Ram PT, Iyengar R (2002) MAP kinase phosphatase as a locus of flexibility in a mitogen-activated protein kinase signaling network. *Science* 297:1018–1023.
2. Hooshangi S, Thiberge S, Weiss R (2005) Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc Natl Acad Sci USA* 102:3581–3586.
3. Cai L, Dalal CK, Elowitz MB (2008) Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature* 455:485–490.
4. Celani A, Vergassola M (2010) Bacterial strategies for chemotaxis response. *Proc Natl Acad Sci USA* 107:1391–1396.
5. Baumgartner BL, et al. (2011) Antagonistic gene transcripts regulate adaptation to new growth environments. *Proc Natl Acad Sci USA* 108:21087–21092.
6. O'Shaughnessy EC, Palani S, Collins JJ, Sarkar CA (2011) Tunable signal processing in synthetic MAP kinase cascades. *Cell* 144:119–131.
7. Walter E, Pronzato L (1997) *Identification of Parametric Models from Experimental Data* (Springer, Berlin).
8. Bennett MR, et al. (2008) Metabolic gene regulation in a dynamically changing environment. *Nature* 454:1119–1122.
9. Hersen P, McClean MN, Mahadevan L, Ramanathan S (2008) Signal processing by the HOG MAP kinase pathway. *Proc Natl Acad Sci USA* 105:7165–7170.
10. Muzzey D, Gómez-Uribe CA, Mettetal JT, van Oudenaarden A (2009) A systems-level analysis of perfect adaptation in yeast osmoregulation. *Cell* 138:160–171.
11. Shimizu TS, Tu Y, Berg HC (2010) A modular gradient-sensing network for chemotaxis in Escherichia coli revealed by responses to time-varying stimuli. *Mol Syst Biol* 6:382.
12. Pelet S, et al. (2011) Transient activation of the HOG MAPK pathway regulates bimodal gene expression. *Science* 332:732–735.
13. Rao CV (2012) Expanding the synthetic biology toolbox: Engineering orthogonal regulators of gene expression. *Curr Opin Biotechnol*, 10.1016/j.copbio.2011.12.015.
14. Voigt CA (2006) Genetic parts to program bacteria. *Curr Opin Biotechnol* 17:548–557.
15. Khalil AS, Collins JJ (2010) Synthetic biology: Applications come of age. *Nat Rev Genet* 11:367–379.
16. Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. *Science* 297:1183–1186.
17. Balázsi G, van Oudenaarden A, Collins JJ (2011) Cellular decision making and biological noise: From microbes to mammals. *Cell* 144:910–925.
18. Munsky B, Neuert G, van Oudenaarden A (2012) Using gene expression noise to understand gene regulation. *Science* 336:183–187.
19. Milias-Argeitis A, et al. (2011) In silico feedback for in vivo regulation of a gene expression circuit. *Nat Biotechnol* 29:1114–1116.
20. de Nadal E, Alepuz PM, Posas F (2002) Dealing with osmostress through MAP kinase activation. *EMBO Rep* 3:735–740.
21. Hohmann S (2002) Osmotic stress signaling and osmoadaptation in yeasts. *Microbiol Mol Biol Rev* 66:300–372.
22. Miermont A, Uhlendorf J, McClean M, Hersen P (2011) The dynamical systems properties of the HOG signaling cascade. *J Signal Transduct* 2011:930940.
23. Yi TM, Huang Y, Simon MI, Doyle J (2000) Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc Natl Acad Sci USA* 97:4649–4653.
24. Ferreira C, et al. (2005) A member of the sugar transporter family, Stl1p is the glycerol/H+ symporter in Saccharomyces cerevisiae. *Mol Biol Cell* 16:2068–2076.
25. O'Rourke SM, Herskowitz I (2004) Unique and redundant roles for HOG MAPK pathway components as revealed by whole-genome expression analysis. *Mol Biol Cell* 15:532–542.
26. Klipp E, Nordlander B, Krüger R, Gennemark P, Hohmann S (2005) Integrative model of the response of yeast to osmotic shock. *Nat Biotechnol* 23:975–982.
27. Hao N, et al. (2007) A systems-biology analysis of feedback inhibition in the Sho1 osmotic-stress-response pathway. *Curr Biol* 17:659–667.
28. Mettetal JT, Muzzey D, Gómez-Uribe C, van Oudenaarden A (2008) The frequency dependence of osmo-adaptation in Saccharomyces cerevisiae. *Science* 319:482–484.
29. Zi Z, Liebermeister W, Klipp E (2010) A quantitative study of the Hog1 MAPK response to fluctuating osmotic stress in Saccharomyces cerevisiae. *PLoS ONE* 5:e9522.
30. Zechner C, et al. (2012) Moment-based inference predicts bimodality in transient gene expression. *Proc Natl Acad Sci USA* 109:8340–8345.
31. Findeisen R, Imsland L, Allgower F, Foss BA (2003) State and output feedback nonlinear model predictive control: An overview. *Eur J Control* 9:179–195.
32. Tamás MJ, et al. (1999) Fps1p controls the accumulation and release of the compatible solute glycerol in yeast osmoregulation. *Mol Microbiol* 31:1087–1104.
33. Csete ME, Doyle JC (2002) Reverse engineering of biological complexity. *Science* 295:1664–1669.
34. Iglesias PA, Ingalls BP (2009) *Control Theory and Systems Biology* (MIT Press, Cambridge, MA).
35. Uhlendorf J, Bottani S, Fages F, Hersen P, Batt G (2011) Towards real-time control of gene expression: Controlling the hog signaling cascade. *Pac Symp Biocomput* 2011:338–349.
36. Menolascina F, di Bernardo M, di Bernardo D (2011) Analysis, design and implementation of a novel scheme for in-vivo control of synthetic gene regulatory network. *Automatica* 47:1265–1270.
37. Toettcher JE, Gong D, Lim WA, Weiner OD (2011) Light-based feedback for controlling intracellular signaling dynamics. *Nat Methods* 8:837–839.
38. Regot S, et al. (2011) Distributed biological computation with multicellular engineered networks. *Nature* 469:207–211.
39. Sprinzak D, et al. (2010) Cis-interactions between Notch and Delta generate mutually exclusive signalling states. *Nature* 465:86–90.
40. Edelstein A, Amodaj N, Hoover K, Vale R, Stuurman N (2010) Computer control of microscopes using μManager. *Curr Protoc Mol Biol*, 10.1002/0471142727.mb1420s92.
41. Rasband WS (1997–2012) *ImageJ* (US National Institutes of Health, Bethesda, MD). Available at http://imagej.nih.gov/ij/. Accessed July 29, 2012.
42. Ballard DH (1981) Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognit* 13(2):111–122.

Supporting Information

Uhlendorf et al. 10.1073/pnas.1206810109

SI Materials and Methods

Mathematical Model. The hyperosmotic stress response is one of the best-studied adaptation processes in yeast (1). Many mathematical models with various levels of complexity have been developed (2–6). In particular, the work by Mettetal et al. (2) showed that essential aspects of the high osmolarity glycerol signaling cascade (amplitude and temporal profile of the protein Hog1 nuclear localization) can be captured by an extremely simple two dimensional linear model. Because simplicity is critical for state estimation (see below), we took inspiration from this minimalistic model, and we propose the following ordinary differential equation (ODE) model:

$$\dot{x}_1 = u(t-\tau) - g_1 x_1$$

and

$$\dot{x}_2 = k_2 x_1 - g_2 \frac{x_2}{K + x_2}.$$

The two variables x_2 and x_1 denote the protein fluorescence level and the recent osmotic stress felt by the cell, respectively, whereas u denotes the osmolarity in the imaging chamber. The recent osmotic stress x_1 is modeled by a term integrating the osmotic stress $[u(t-\tau)]$ and a linear decay term $(-g_1 x_1)$, which lets the influence of past osmotic stresses diminish exponentially with time. The term describing the osmolarity (u) is computed based on (a piecewise linearization of) the profile shown in Fig. S1 and the valve status. A delay in u aggregates times for signal transduction, gene expression, and protein synthesis. The fluorescence level increases linearly with the integrator with rate k_2. Because we did not observe any saturation of the protein fluorescence levels, even for long time experiments (>10 h), in characterization experiments (Fig. S2), we deduced that protein degradation should not be modeled as a term proportional to its concentration as usually assumed. A good fit altogether was obtained using a saturating term for its degradation rate. To fit our model parameters, we designed two types of experiments to probe different aspects of the cell response dynamics: repeated or isolated osmotic shocks. For both cases, we performed experiments with shocks lasting 5, 6, 7, or 8 min (Fig. S2). Based on these observations, we manually set the delay τ to 20 min and the protein level corresponding to degradation half-saturation K to 750. In these conditions, degradation does not saturate in single shocks experiments. The three other parameters have been set using the global optimization tool CMAES (7) so that they minimize the mean squared deviation between model predictions and observations over all experiments. All parameters are summarized in Table S1. Despite its extreme simplicity, we were able to obtain an acceptable fit to the data (Fig. 1 and Fig. S2).

State Estimation. For linear systems, the standard approach for state estimation is to use a Kalman filter. Using a linear model, the past control inputs, and (noisy) measurements, it gives an estimate of the system state together with its uncertainty. For each estimate, the filter chooses a weighted average between observation and model prediction, the weight of each term depending on its associated uncertainty. Kalman filters can be extended to nonlinear systems (EKF) by linearizing the nonlinear model around the operating points (8). Note that the delay in the model only concerns the input. In consequence, a classical EKF can be used without the need for specific delayed system state estimation techniques. An EKF requires the tuning of few parameters, the covariance matrix of the measurement error (R), and the covariance matrix of the process noise (Q), where errors in both cases are assumed to be Gaussian-distributed with zero mean. Although it is very difficult to evaluate the quality of the state estimation process, we were able to obtain good control performances after setting R to 2,500 and Q to diag(0.37, 925). The value of R has been chosen by estimating the variance of the observations from experimental data. The ratio of the diagonal values in Q has been determined by the steady-state values of x_1 and x_2. Then, the actual values of Q have been determined by ensuring consistency between the innovation residuals (the difference between model prediction and observation) and the variance of the innovation residuals, which is estimated by the Kalman filter.

Comparison Between Single-Cell Control and Population Control Quality. To compare the control quality of the control loop across different experiments, we consider the mean square deviation (MSD) of the measured signal with respect to the target profile. For single-cell control experiments, the signal is simply the fluorescence of the controlled cell. For population control experiments, signals can either be the mean of the fluorescence across the population of cells or the fluorescence of any cell within the controlled population. In the first case, we refer to the MSD of the mean, and in the second case, we refer to single-cell MSDs in a population experiment. The mean MSD is then the mean of the single-cell MSDs. Note the difference between the MSD of the mean and the mean MSD (Fig. 4 and Fig. S6). In set-point and sine wave experiments, only the time period for which the cells effectively follow the target has been considered (the first 250 min of the initialization phase were disregarded). To allow for a proper quantification of control performance, only cells that have been tracked continuously for at least 5 (set point) or 7 (tracking) hours were considered.

For set-point experiments, in which the fluorescence of the cells remained around a fixed target value, we defined the noise level of a measured signal as its SD over its mean. As for control quality, we compared (in Fig. S6) the single-cell noise level in single-cell and population control experiments with the noise level of the mean and the mean noise level for population control experiments.

We wondered whether controlling cells individually gives statistically better results than controlling them as members of a population. The null assumption (H_0) was, therefore, that MSDs (or noise levels) of cells in single-cell control experiments or population control experiments follow the same distribution. The alternative assumption (H_1) was that the MSDs (or noise levels) in single-cell control experiments are statistically smaller than in population control experiments. Because we worked with samples of small sizes and could not a priori assume that MSDs or noise levels follow a known probability distribution, we used a nonparametric method (9), the Fligner–Policello (FP) test (10). Like the Wilcoxon–Mann–Whitney test, the FP test is one of the most powerful nonparametric tests used to check whether two independent samples have been drawn from the same distribution. However, in contrast to the Wilcoxon–Mann–Whitney test, the FP test allows for comparison of samples having different variances (Behrens–Fisher problem) (9). This difference is important, because we expected that single-cell control reduced the variance of MSDs. In its one-sided form, the FP test addresses precisely the question stated above (H_0 vs. H_1). As shown in Table S2, the test revealed that single-cell control does statistically improve control performances for both MSDs (at least in one case) and noise levels.

1. Hohmann S, Mager WH (2003) *Yeast Stress Responses* (Springer, Berlin).
2. Mettetal JT, Muzzey D, Gómez-Uribe C, van Oudenaarden A (2008) The frequency dependence of osmo-adaptation in Saccharomyces cerevisiae. *Science* 319:482–484.
3. Muzzey D, Gómez-Uribe CA, Mettetal JT, van Oudenaarden A (2009) A systems-level analysis of perfect adaptation in yeast osmoregulation. *Cell* 138:160–171.
4. Zi Z, Liebermeister W, Klipp E (2010) A quantitative study of the Hog1 MAPK response to fluctuating osmotic stress in Saccharomyces cerevisiae. *PLoS One* 5:e9522.
5. Klipp E, Nordlander B, Krüger R, Gennemark P, Hohmann S (2005) Integrative model of the response of yeast to osmotic shock. *Nat Biotechnol* 23:975–982.
6. Hao N, et al. (2007) A systems-biology analysis of feedback inhibition in the Sho1 osmotic-stress-response pathway. *Curr Biol* 17:659–667.
7. Hansen N, Ostermeier A (2001) Completely derandomized self-adaptation in evolution strategies. *Evol Comput* 9:159–195.
8. Jazwinski AH (1970) *Stochastic Processes and Filtering Theory*, ed Bellman R (Academic, London).
9. Siegel S, Castellan NJ (1988) *Nonparametric Statistics for the Behavioral Sciences* (McGraw-Hill, New York).
10. Fligner MA, Policello GE (1981) Robust rank procedures for the Behrens-Fisher problem. *J Am Stat Assoc* 76:162–168.

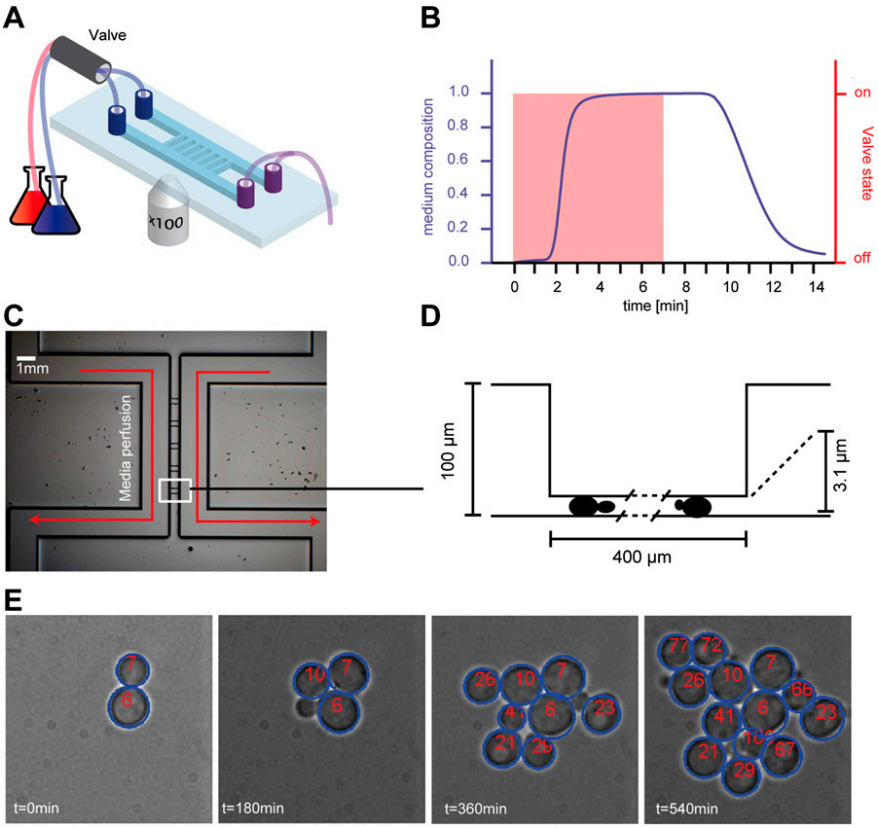


Fig. S1. The microfluidic device. (*A*) Schematic representation of the microfluidic device. A flow is created in the channels thanks to a peristaltic pump placed downstream. Upstream, a valve allows switching between the two media. (*B*) Using ink, we measured the dynamics of fluid exchange. This switching profile was obtained in a robust manner. (*C*) Close up of the microfluidic device. (*D*) Cells are captured within thin chambers. The media in the imaging chamber are exchanged by diffusion. (*E*) An example of cell imaging, segmentation, and tracking is shown.

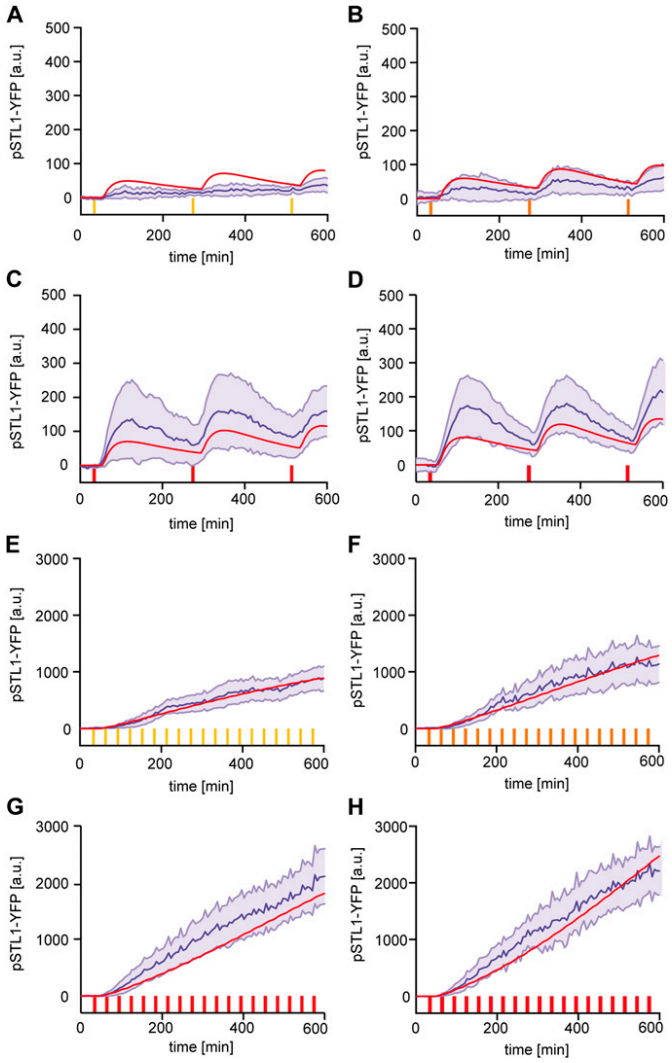


Fig. S2. Calibration experiments. (*A–D*) Isolated osmotic shocks lasting 5, 6, 7, and 8 min were applied every 4 h. (*E–H*) Osmotic shocks lasting 5, 6, 7 and 8 min were applied every 30 min. The fluorescence of yECitrine, controlled by the *STL1* promoter, was quantified. Experimental means and SDs are represented as blue lines and blue envelopes. Model predictions using the parameters of Table S1 are represented in red. Note that the scales for fluorescence intensities in repeated or isolated shock experiments are significantly different.

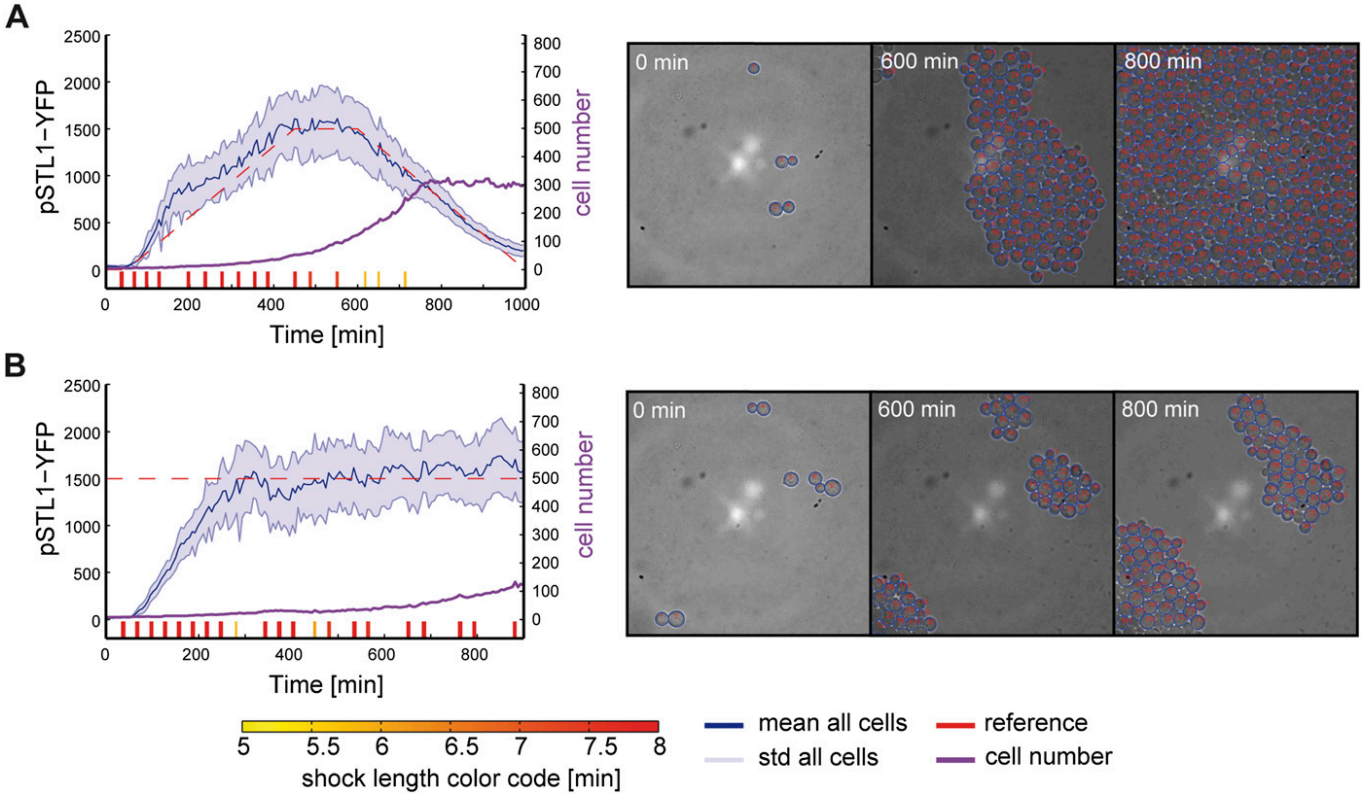


Fig. S3. Cell growth during experiments. (*A*) Cell growth in the field of view. The number of cells detected in the field of view is depicted in violet. In most experiments, the field of view is completely covered with cells after 800–900 min. Cells growing out of the imaging chamber are washed away by the flow, and the number of cells detected in the field of view reaches a plateau (the maximal number of cells in the field of view is around 300). (*B*) Strong osmotic shocks are known to transiently stop the cell cycle. In this experiment, repeated strong osmotic inputs were applied to reach the target value of 1,500 fluorescence units (f.u.). These inputs led to a slowdown of the cell cycle, and the cells grew slower compared with the experiment shown in *A*.

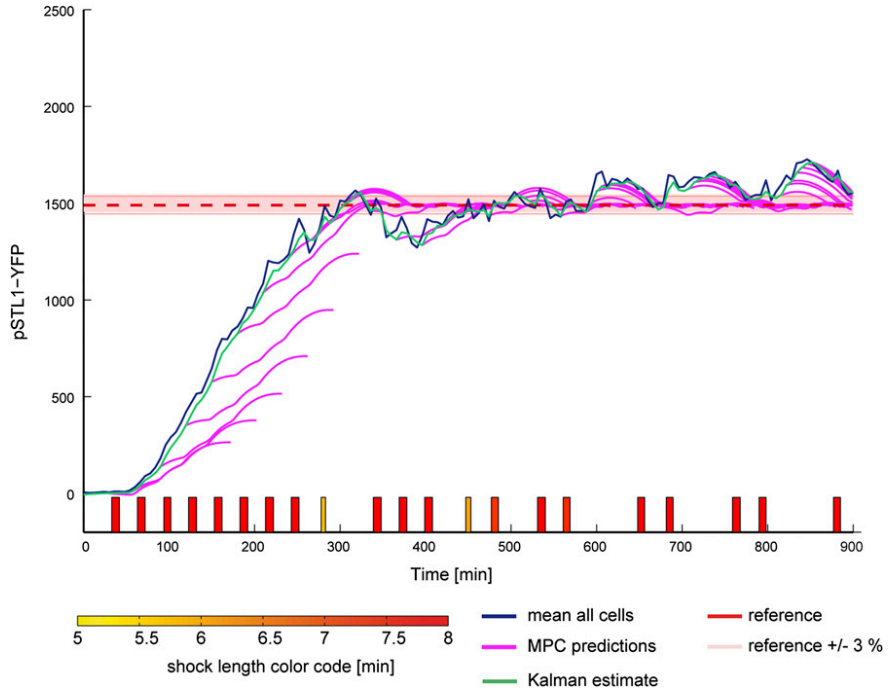


Fig. S4. Performance of the model predictive control algorithm. In this plot of the population control experiments from Fig. 2*B*, the Kalman estimation is indicated as a green line, and the predictions of the best control strategy found at each iteration of the model predictive controller are plotted as violet lines. It shows that good (theoretical) control strategies were found by the controller. Limitations are, therefore, likely to come essentially from model inaccuracies and biological variability. The red envelope around the reference indicates a region close to the reference (±3%).

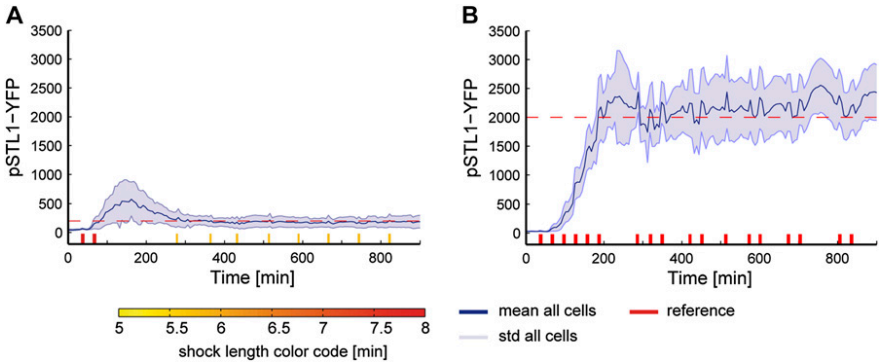


Fig. S5. Assessing control capabilities of the proposed platform. We investigated the lower and higher limits for set-point control using the proposed platform. (*A*) In the first experiment, we set the target value at a very low level: 200 f.u. The control result shows a significant overshoot at the beginning, but after ~300 min, the observed fluorescence level follows the target value faithfully. Lower control objectives are hardly possible with our experimental setting, because these control targets would be within the background fluorescence level of a cell (around 30 f.u.). (*B*) In the second experiment, we tested the upper limits of the control platform by setting the target value at 2,000 f.u. The control works but shows significant levels of noise. These results show that one can control gene expression within a 10-fold range using the proposed experimental setting.

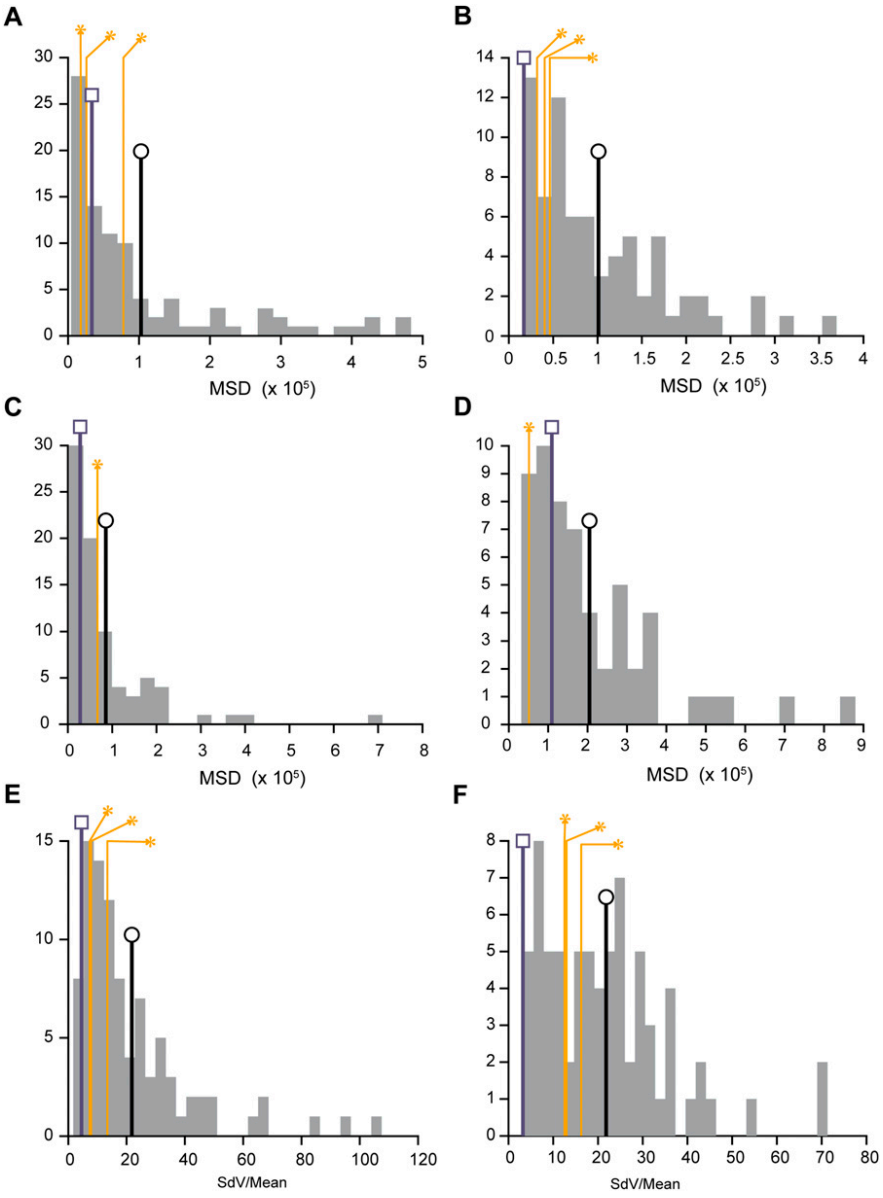


Fig. S6. Comparison of the control qualities and the noise levels between single-cell and population control experiments. (*A–D*) Histograms showing the distribution of single-cell MSDs in population control experiments. MSD of the mean (blue bar, square), mean MSD (black bar, circle), and MSD of the controlled cells in three single-cell control experiments (orange, star). (*A*) Target control at 1,000 f.u. (*B*) Target control at 1,500 f.u. (*C*) Trapeze control (Fig. 2). (*D*) Sine wave control (Fig. 2). (*E* and *F*) Histograms showing the noise levels of single-cell fluorescence in population control experiments. Noise level of the mean (blue bar, square), mean noise level (black bar, circle), and noise levels of the controlled cell in three single-cell control experiments (orange, star). *E* and *F* correspond to set-point control experiments at a value 1,000 and 1,500, respectively.

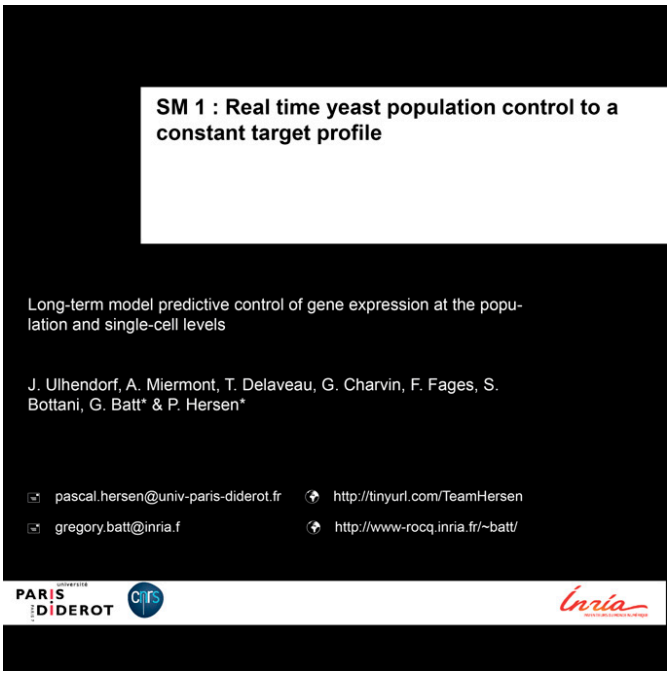
Table S1. List of parameter values

Name	Value
g_1	$4.02\ 10^{-3}$
g_2	37.5
τ	20
k_2	0.581
K	750

Table S2. Computation of the Ù statistic, its associated *P* value for MSDs and noise levels, and the two set-point control experiments using the FP test

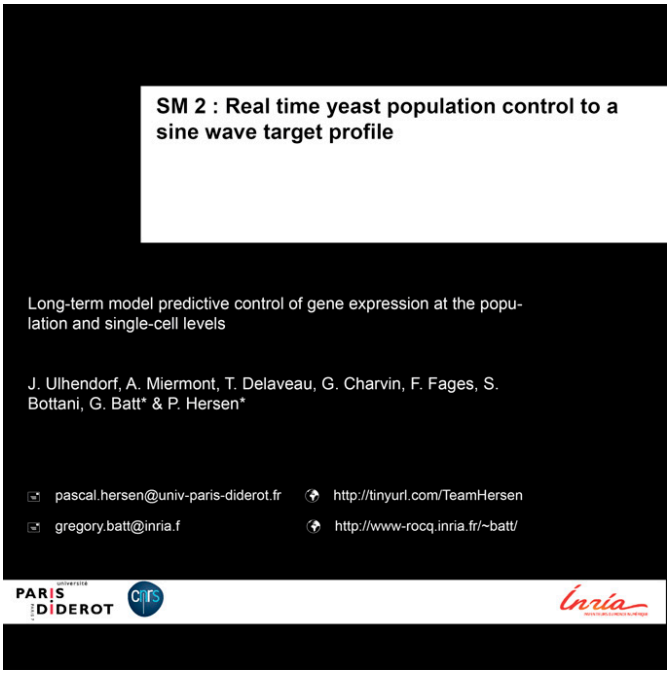
Experiment	N1	N2	Ù	*P* value*
MSD, T = 1,000 f.u. (Fig. S6*A*)	92	3	171	0.15
MSD, T = 1,500 f.u. (Fig. S6*B*)	73	3	172	$1.7\ 10^{-7}$
Noise level, T = 1,000 f.u. (Fig. S6*E*)	92	3	202	$3.7\ 10^{-3}$
Noise level, T = 1,500 f.u. (Fig. S6*F*)	73	3	144	$8.5\ 10^{-3}$

N1 and N2 are sample sizes.
*As standard for the FP test, the *P* value computation uses the normal approximation.



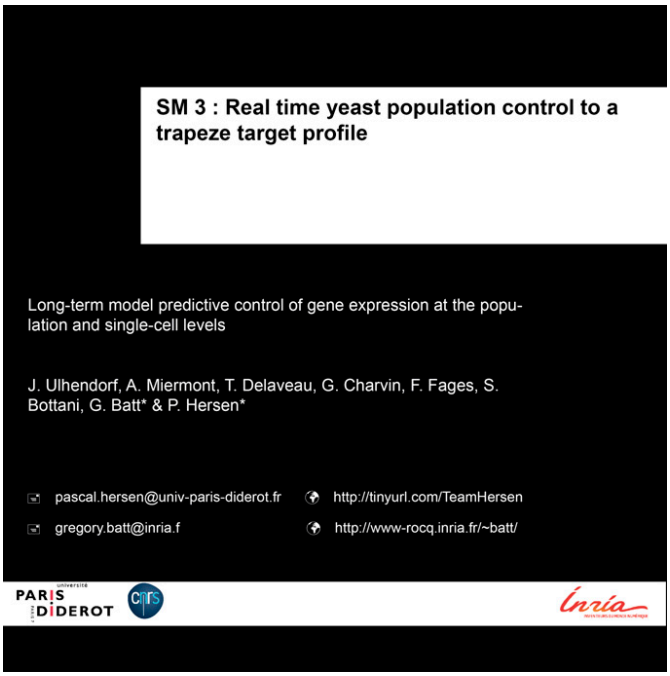
Movie S1. Population control I. The target profile is constant (T = 1,500 f.u.).

Movie S1



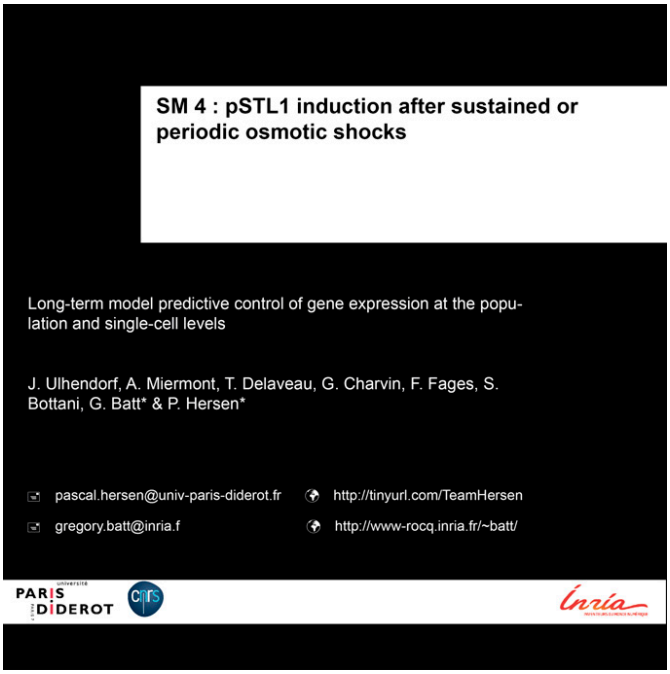
Movie S2. Population control II. The target profile is a sine wave.

Movie S2



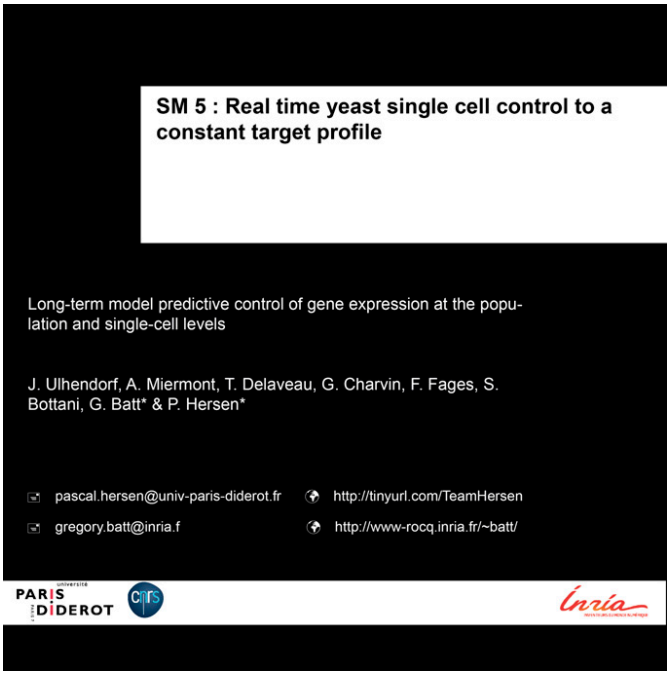
Movie S3. Population control III. The target profile is a trapeze.

Movie S3



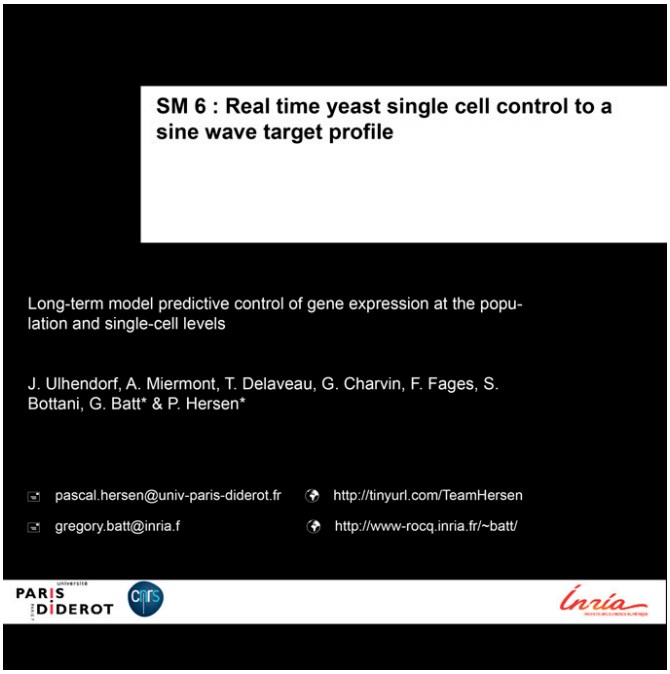
Movie S4. Response to sustained or periodic hyperosmotic shock (1 M sorbitol).

Movie S4



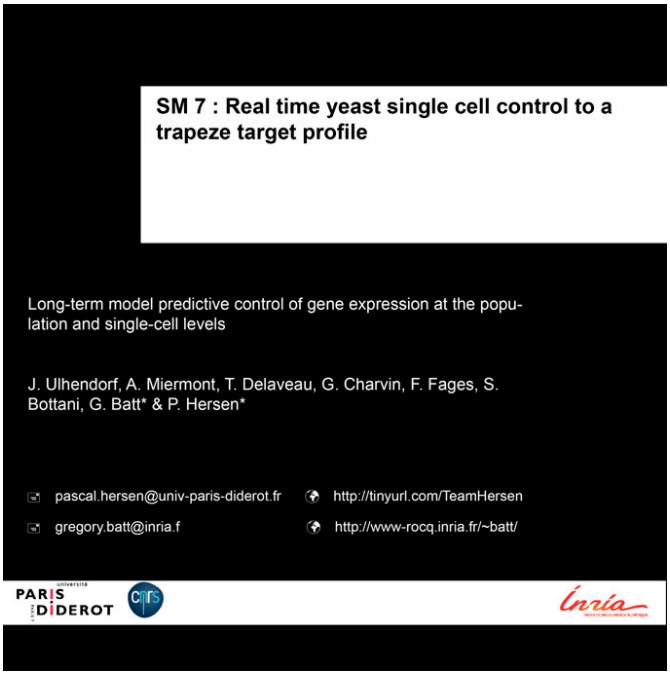
Movie S5. Single-cell control I. The target profile is constant (T = 1,500 f.u).

Movie S5



Movie S6. Single-cell control II. The target profile is a sine wave.

Movie S6



Movie S7. Single-cell control III. The target profile is a trapeze.

Movie S7

B.4 Appended paper 4: The dynamical systems properties of the HOG signaling cascade

Review: The dynamical systems properties of the HOG signaling cascade

Agnès Miermont, Jannis Uhlendorf, Megan McClean and Pascal Hersen

Journal of Signal Transduction, 2011

Hindawi Publishing Corporation
Journal of Signal Transduction
Volume 2011, Article ID 930940, 12 pages
doi:10.1155/2011/930940

Review Article

The Dynamical Systems Properties of the HOG Signaling Cascade

Agnès Miermont,[1] Jannis Uhlendorf,[1,2] Megan McClean,[3] and Pascal Hersen[1,4]

[1] *Laboratoire Matière et Systèmes Complexes, UMR7057, CNRS and Université Paris Diderot, 10 rue Alice Domon et Léonie Duquet, 75013 Paris, France*
[2] *Contraintes Research Group, Institut National de Recherche en Informatique et en Automatique, INRIA Paris—Rocquencourt, France*
[3] *Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08540, USA*
[4] *The Mechanobiology Institute, National University of Singapore, 5A Engineering Drive 1, Singapore 117411, Singapore*

Correspondence should be addressed to Pascal Hersen, pascal.hersen@univ-paris-diderot.fr

Received 14 August 2010; Revised 19 October 2010; Accepted 12 November 2010

Academic Editor: Johannes Boonstra

The High Osmolarity Glycerol (HOG) MAP kinase pathway in the budding yeast *Saccharomyces cerevisiae* is one of the best characterized model signaling pathways. The pathway processes external signals of increased osmolarity into appropriate physiological responses within the yeast cell. Recent advances in microfluidic technology coupled with quantitative modeling, and techniques from reverse systems engineering have allowed yet further insight into this already well-understood pathway. These new techniques are essential for understanding the dynamical processes at play when cells process external stimuli into biological responses. They are widely applicable to other signaling pathways of interest. Here, we review the recent advances brought by these approaches in the context of understanding the dynamics of the HOG pathway signaling.

1. Introduction

Living organisms have evolved specialized biochemical pathways to cope with stressful, often changing environments. Even in simple cells such as yeast, thousands of specialized sets of sensing and signaling proteins form modules used to monitor and adapt to the environmental state and its variations. Such modules can be insulated or, on the contrary, connected to one another. Whereas insulation allows for robust and sensitive response, the interconnection of modules allows for higher-level behavior such as multiple input sensing and decision making through cross-talk [1]. For a given stimulus, the biochemical components of the different modules that play a role in the cellular response are usually well described in the literature. Their biological functions and interactions are known in detail, especially in model organisms such as the budding yeast. This knowledge comes from decades of complex, tedious, and elegant experiments. Genetic techniques such as gene deletion, mutation, and overexpression have been used to infer the connection patterns between proteins and the architectures of many modular functions. Biochemical assays provided crucial information on protein phosphorylation and kinase activity. Microarrays revealed the role of these modules in determining global gene expression.

Signaling pathways are naturally dynamic [2] in that cells must respond to external signals in a timely manner, and indeed, the cellular response is often affected by the temporal properties of the external signal. In addition, the internal dynamics and timing of events in the signaling pathway determine the cellular response. These internal dynamics determine the information flow, allowing cells to process and convey information from a sensory input to a specific protein in charge of orchestrating the cellular response [3]. Until recently, experimental techniques have been limited such that most studies have examined the response of a signaling pathway to a stationary stimulus. Accordingly, adaptation and cellular responses to environmental cues were usually studied only with respect to the magnitude of the stimulus without seriously taking into account dynamical aspects. Identification of the components of a signaling pathway through the techniques mentioned above, combined with studies of simple stationary stimuli, is not enough to

understand the dynamics or systems-level properties of a complex biological network.

With the emergence of systems biology, there has been an important paradigm shift, and it is becoming increasingly clear that the temporal variations of stimulatory inputs can be directly sensed by cells [5] and that studying cells in time-variable environments is a powerful way to determine signaling pathway architecture and to understand how they process information [6, 7]. Experimental microfluidics-based strategies have matured to allow for excellent control of the cellular environment both in time and space [8, 9]. This technology coupled with genetic engineering to fluorescently tagged proteins allows for real-time observation of the system's response using fluorescence microscopy. Finally, quantitative real-time measurements form the basis for the development of mathematical models and the use of signal analysis tools, such as reverse engineering, to model the dynamical aspects of signaling pathways [10]. These models in turn provide testable experimental predictions.

This review describes the recent strategies that have been developed to assess quantitatively the dynamics of the canonical HOG MAP kinase (MAPK) pathway in the yeast, *Saccharomyces cerevisiae*. We shall first briefly review the key characteristics of the organization of the HOG pathway. We then discuss the novel experimental and modeling tools [10, 11, 16–18] that are allowing new insights into the pathway's dynamics and systems-level behavior.

2. MAPK Cascades in Yeast

Among signaling pathways, the Mitogen Activated Protein Kinases (MAPK) family has received considerable attention. MAPK pathways are very well conserved from yeasts to mammals [19–21] and several comprehensive reviews are available in the recent literature [22, 23]. MAP Kinase pathways are involved in many cellular processes such as stress response, the regulation of differentiation and proliferation. These pathways contain a canonical module of three protein kinases that act in series (Figure 1). Upon phosphorylation by an upstream protein, a MAP kinase kinase kinase (MAPKKK) phosphorylates a MAP kinase kinase (MAPKK) on conserved serine and threonine residues, which in turn phosphorylates a MAP kinase (MAPK) on a threonine (sometimes serine) and a tyrosine residue located adjacent to each other and separated by a single amino acid (Thr/Ser-X-Tyr). This dual phosphorylation site is located in the activation loop of the catalytic domain and its dual phosphorylation is needed for activation of the MAP kinase.

There are five MAPK modules in yeast (Table 1) [22]. The hyperosmotic glycerol (HOG) pathway is activated in response to a hyperosmotic stress [24–28]. The Cell Wall Integrity (CWI) module controls the cell wall integrity and is triggered in response to numerous stresses including cell wall deterioration, temperature shifts, and hypo osmotic shocks [29–31]. The pheromone pathway [32, 33] controls the mating response which involves an important morphological deformation of yeast cells. Finally, the filamentous growth pathway [33, 34] and the sporulation pathway [22] control

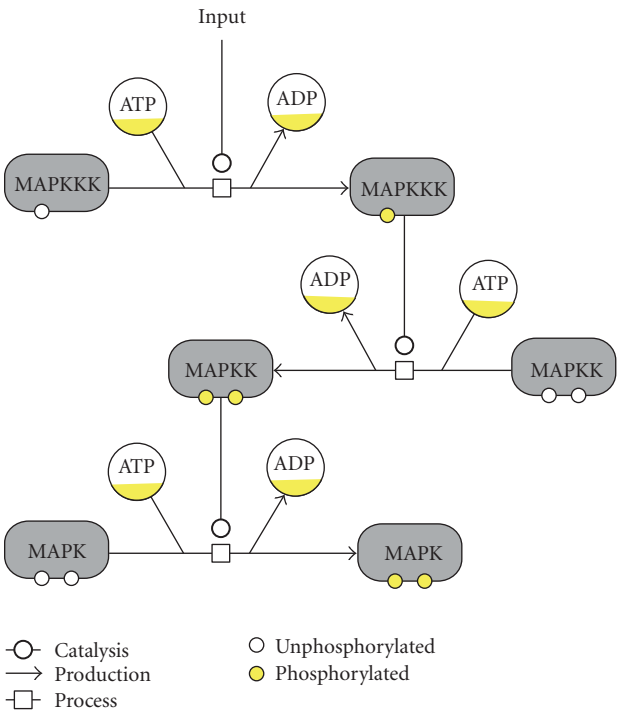


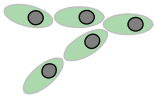
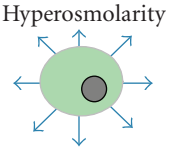
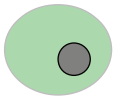
Figure 1: The canonical structure of a MAPK cascade. We used the Systems Biology Graphical Notation (SBGN) [4] to represent the interactions between the MAP Kinases. Activations of MAPK occur through enzymatic phosphorylation and ATP consumption. Interactions with other components and in particular with phosphatases are not shown. In the case of the HOG pathway in yeast, dual phosphorylation of the final MAPK (Hog1p) occurs within a few minutes after an hyper-osmotic stimulus.

the response to starvation for haploid and diploid cells although the sporulation pathway is not as well known as the other four MAPK pathways. Only its MAPK has been identified in diploid cells (Smk1p), and it is thought to drive the spore cell wall assembly [22]. Though they share numerous components, the five MAPK pathways of the yeast *Saccharomyces cerevisiae* are tightly regulated by cross-talk and mutual inhibition which permit faithful signaling, adaptation to their environment, and regulation of growth and morphogenesis [22]. Among these MAPK pathways, the HOG pathway (Figure 2) is particularly well suited to study signaling dynamics, since it can be reliably activated through increasing the osmolarity of the environment.

3. The HOG MAPK Signaling Pathway

Water homeostasis is fundamental for life. In nature, the environment can vary rapidly from isotonic to hyper or hypo osmotic conditions, and yeast cells have to adapt quickly [23, 47]. The first response after a hyperosmotic shock is the rapid loss in few seconds of cell volume due to water efflux and the activation of membrane sensory receptors followed by the activation of the HOG pathway which is completed after a few minutes (Figure 2) [11, 48]. Two distinct branches of the pathway detect changes in osmolarity and activate the

Table 1: The MAPK pathways in *S. cerevisiae*. The morphological adaptation corresponds to the cell behavior in response to each specific signaling input. The major molecular actors for each pathway are indicated below. Spore cell wall assembly during sporulation is another morphogenetic process driven by a MAPK protein (Smk1p), but with little knowledge on the other proteins involved and the structure of the pathway.

External Stress	Pheromone	Starvation	Hyperosmolarity	Cell wall Stress
Morphological Adaptation				
Membrane Sensors	Ste2/3	Sho1	Sln1, Sho1, Msb2, Hkr1, Opy2	Wsc1, Mid2
MAPKKK	Ste11	Ste11	Ssk2/22 & Ste11	Bck1
MAPKK	Ste7	Ste7	Pbs2	Mkk1/Mkk2
MAPK	Fus3/Kss1	Kss1	Hog1	Slt2
Transcription factors	Ste12 [22, 35]	Ste12, Tec1 [22, 35, 36]	Hot1, Sko1, Smp1, Msn2/4 [22, 28, 37, 38]	Rlm1, Swi4/6 [22, 39, 40]
Inhibition	Msg5, Ptp2, Ptp3 [21, 41, 42]	—— [21]	Ptcs, Ptp2, Ptp3 [21, 43–45]	Msg5, Ptp2, Ptp3, Sdp1 [21, 30, 46]

pathway. These branches converge at the level of the MAPKK Pbs2p. The first branch is referred to as the SHO1 branch [23, 49], while the second is referred to as the SLN1 branch [50, 51].

Sln1p negatively regulates the HOG signaling pathway and deletion of *SLN1* is lethal due to pathway overactivation. This lethality is suppressed by knocking out any of the downstream components *SSK1*, *SSK2/SSK22*, *PBS2*, or *HOG1*. Sln1p contains two transmembrane domains, a histidine kinase domain and a receiver sequence. Sln1p autophosphorylates on its histidine kinase domain. The phosphate group is then transferred to its receiver domain, then to Ypd1p and finally to Ssk1p. This set of three proteins forms a phosphorelay [50, 51], a very common signaling motif in prokaryotes [51], but rare in eukaryotic cells such as yeast. The phosphorylated form of Ssk1p is inactive and the downstream MAPK pathway is usually not activated. However, after a hyperosmotic stress, Sln1p is inactivated by an unknown mechanism (though it has been proposed that Sln1p is sensitive to membrane tension [23, 52]) leading to the inactivation of Ypd1p and derepression of Ssk1p. Finally, unphosphorylated Ssk1p binds to the MAPKKKs Ssk2p and Ssk22p, which autophosphorylate, and then can phosphorylate the MAPKK Pbs2p. Sln1p seems to dominate the kinetic response of the pathway while also ensuring its robustness by inducing high basal Hog1p expression counteracted by a fast-acting negative feedback to allow rapid pathway response [53]. Thus, this tightly tuned signaling branch allows rapid and sensitive responses to environmental changes.

Sho1p consists of four transmembrane domains and an SH3 domain. This domain permits the recruitment of molecular actors, notably the MAPKK Pbs2p, to the plasma membrane [54]. The upstream kinase Ste20p, the G-protein Cdc42p, and the MAPKKK Ste11p needed for the activation of the protein Pbs2p are also recruited to the membrane [55]. Since it is a transmembrane protein, Sho1p has long been considered an osmosensor [56]. However, recent studies suggest that Sho1p is more an anchor protein than a sensor for osmolarity [55]. Hkr1p and Msb2p, two mucin-like [57–60] proteins that form heterooligomeric complexes with Sho1p [58, 59] have recently been proposed as osmosensors of the SHO1 branch. Components of the SHO1 branch also take part in pseudohyphal development and mating, indicating that Sho1p might not have a specific role in osmosensing but a more general role related to cell shape measurement [61].

MAPKKKs of these two initiating branches induce the phosphorylation of the MAP kinase kinase Pbs2p on the conserved residues Ser514 and Thr518 [62]. Pbs2p is a cytoplasmic protein essential for the activation of Hog1p by dual phosphorylation on the conserved Thr174 and Tyr176 [62]. *PBS2* and *HOG1* are essential for osmoadaptation as null mutations in both genes induce osmosensitivity [23, 63]. Pbs2p also plays the role of a scaffold for the SHO1 branch [49, 54, 56, 64] by anchoring the different components, promoting signal propagation between proper protein partners and preventing improper cross-talk between the Pheromone pathway and the HOG pathway. Once Pbs2p phosphorylates Hog1p, Hog1PP translocates to the nucleus in a manner that is dependent upon the karyopherin Nmd5p [65]. Localization of Hog1p-GFP to the nucleus can be used as a reliable reporter of pathway activity.

4. Sequential Response after a Hyperosmotic Shock

The activation of the Hog1p MAPK triggers several responses on different time-scales (Figure 3) [48]. A rapid non-transcriptional response in the cytoplasm corresponds to the closure of Fps1p [66] and the activation of several kinases (e.g., Rck2p [67], Pfk2p [68]). Fps1p belongs to the ubiquitous Major Intrinsic Protein (MIP) [69] family and is known to play a central role in yeast osmoadaptation by controlling both uptake and efflux of the osmolyte

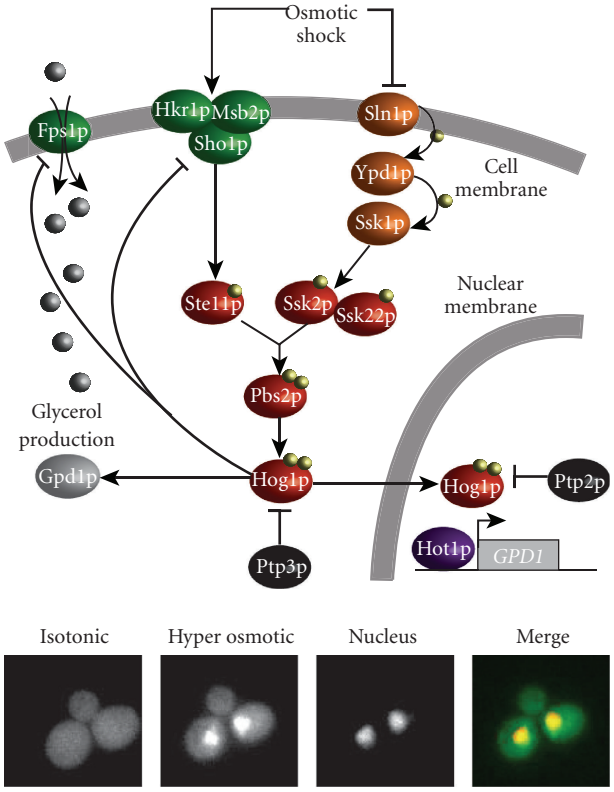


Figure 2: The HOG pathway. View of the main molecular actors involved in the hyperosmotic glycerol pathway (see text for more details). Two branches led by Sho1p and Sln1p are sensitive to high osmolarity and lead to the activation of Pbs2p and Hog1p after a hyperosmotic shock. Hog1p has both a cytoplasmic and a nuclear role, with different timescales, that correspond to a fast non transcriptional response and a longer response involving transcription when dealing with strong hyper osmotic shock. The yeast pictures at the bottom show nuclear localization of Hog1p tagged by GFP after a moderate hyper-osmotic shock (Sorbitol, 1 M). Colocalization with the nucleus is seen on the overlay pictures between the GFP channel (Hog1p) and the RFP channel (Htb2p). Note that localization is transient and reversible if the cell is put back into isotonic conditions.

glycerol [70]. Importantly, Fps1p is gated by osmotic changes [66, 71]. Indeed, this channel protein is closed under hyperosmotic stress to enable intracellular accumulation of glycerol, whereas it is open under low-osmolarity conditions to allow for glycerol efflux.

On a longer time scale, several minutes after an osmotic shock, Hog1p induces the modification of expression of nearly 600 genes [72–75]. This transcriptional response is driven by intermediate transcriptional factors: Hot1p, Sko1p, Smp1p, and Msn2/4p [37, 38, 74, 76, 77]. Importantly, Hog1p initiates glycerol biosynthesis via the transcriptional factor Hot1p [38]. Glycerol production is due to the expression of glycerol-3-phosphate dehydrogenase and glycerol-3-phosphatase. Both enzymes are encoded by two similar isogenes, *GPD1*, *GPD2* and *GPP1*, *GPP2*, respectively, [78, 79]. The accumulation of glycerol results in an increase of the internal osmolarity, leading to water influx and cell size recovery. Hot1p is also involved in regulating glycerol influx by inducing a strong and transient expression of *STL1*, which codes for a glycerol proton symporter located in the plasma membrane [80]. Hog1p is dephosphorylated and exported from the nucleus via the karyopherin Xpo1p [65] 20 to 30 minutes after an osmotic shock depending on the severity of the shock. This is concomitant with the onset of glycerol production and restoration of osmotic balance. Dephosphorylation of Hog1p is due to nuclear phosphatases. Phosphatases have a critical role in down-regulation of MAPK proteins whose excessive activation can be lethal for the cell. In yeast, three classes of protein phosphatases are known to downregulate MAPK pathways. The dual specificity phosphatases (DSPs) dephosphorylate both phosphotyrosine (pY) and phosphothreonine (pT). The protein tyrosine phosphatases (PTPs) dephosphorylate only tyrosine residues. Finally, protein phosphatases type 2C (PTC) dephosphorylate threonine, serine, and sometimes tyrosine residues. For the HOG pathway, the serine-threonine phosphatases Ptc1p, Ptc2p, and Ptc3p act on both the Pbs2p (MAPKK) and Hog1p (MAPK), while the tyrosine phosphatases, Ptp2p and Ptp3p strictly control Hog1p [43, 44, 49]. Ptp2p is predominantly localized in the nucleus, Ptp3p in the cytoplasm, while the protein phosphatases types 2C are located both in the cytoplasm and in the nucleus. Simultaneous knockout of both *PTP2* and *PTC1* is lethal for the cell [45]. Deletion of *PTP3* induces overactivation of Hog1p but is not lethal because it predominantly acts on other MAPK proteins involved in the mating pathway.

5. Towards a Model of the HOG Pathway

Years of genetic and biochemical analysis have provided us with an extraordinarily precise description of the key players in the HOG pathway. What about the signaling dynamics of the pathway? How does the architecture determine the pathway's signal processing ability? Classic molecular biology experiments were based on step shock experiments with an osmotic agent, such as NaCl or sorbitol at various concentrations. Phosphorylation states of key proteins have been measured at different time points after a step shock at the population level, showing a transient increase of phosphorylation (lasting several minutes) concomitant with nuclear enrichment of Hog1p [81]. Nuclear cytoplasmic shuttling of Hog1p was also observed qualitatively, indicating a fast deactivation of the pathway when cells are returned to an isotonic environment [81]. Levels of gene expression have been measured at different timepoints after an osmotic shock using microarrays [73]. Although done with a low resolution in time compared to biophysical experiments, these measurements give an idea of the dynamics of the activation of the pathway.

Based on such measurements, several models have been proposed to describe mathematically the HOG signaling pathway and more generally osmoadaptation in yeast [82]. The most comprehensive and the first integrative one is due to Klipp et al. [81]. Their model takes not only the HOG signaling cascade into account (only the SLN1 branch), but

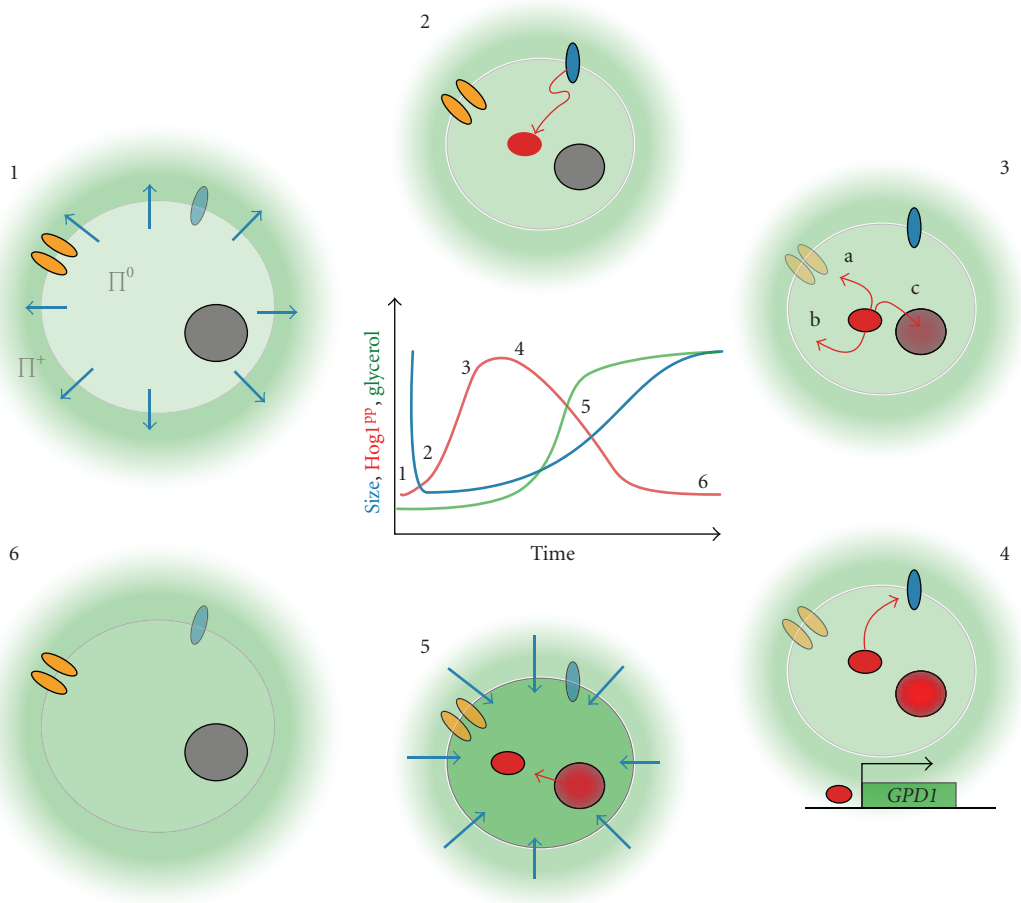


Figure 3: Sequential sketch of yeast adaptation to a hyperosmotic shock. The evolution with time of the size, phosphorylation of Hog1p, and internal concentration of glycerol are schematically represented in the center of the picture. (1) After an increase of the external osmolarity (green), a first mechanical response corresponds to a rapid loss of water (blue arrow). It leads to a decrease of the cell size and a loss of turgor pressure. (2) HOG osmosensors (blue) activate the pathway and eventually lead to the phosphorylation of Hog1p. (3) Hog1PP induces several processes: (a) Inactivation of the glycerol channel Fps1p preventing glycerol leakage; (b) direct or indirect activation of cytoplasmic actors, for example, 6-phosphofructo-2-kinase (Pfk2p) involved in glycerol synthesis; (c) translocation in the nucleus. Note that there are other targets of Hog1p such as Sic1p, Hsl1p, Nha1p, and Tok1p. (4) Nuclear Hog1PP induces a large transcriptional response. In particular, the gene GPD1 leading to glycerol synthesis is upregulated. Negative feedbacks (glycerol production, phosphorylation of Sho1p, etc.) allow inhibition of the pathway activity. (5) Increase of the internal glycerol leads to water influx and progressive cell size recovery while Hog1p is exported from the nucleus. (6) Pathway is off, and turgor pressure and cell size are restored. The cell is adapted to its new environment.

also includes a description for the metabolic production of glycerol, as well as an elementary gene expression model for the enzymes involved in glycerol production. The model also includes the closure of the membrane glycerol tranporter Fps1p and takes the dephosphorylation of nuclear Hog1p by Ptp2p into account. Most reactions in the model were described by the mass action rate law. The model consisted of 70 parameters, of which 24 had to be estimated. To estimate this number of parameters with the limited data available, the authors divided the model in modules and fitted them separately to data points. Their model reproduced accurately the transient response of the HOG pathway after a single hyperosmotic shock. This included the phosphorylation states of Hog1p and Pbs2p, as well as glycerol production and cell-size recovery. In addition, the model was able to correctly predict the effect of different mutations of proteins involved in the pathway. Mutants unable to produce glycerol (*gpd1*Δ, *gpd2*Δ) [83] or to close the Fps1p channel showed an increased duration of HOG activity. Mutants with an increased phosphatase Ptp2p activity showed a lower level of phosphorylated Hog1p but a similar period of HOG activity.

Although very promising, such an approach is still extremely difficult to fine tune since it relies on many unknown parameters. Comparison of the model outputs to experimental data is crucial. To further constrain and test complex models one needs quantitative, time-resolved experiments at the single-cell level in response to complex input signals.

As engineers do with electronic circuits and chips, a very powerful way to explore the dynamics of a given system is to observe its response to complex input signals. Such an approach lends itself to developing minimal models that

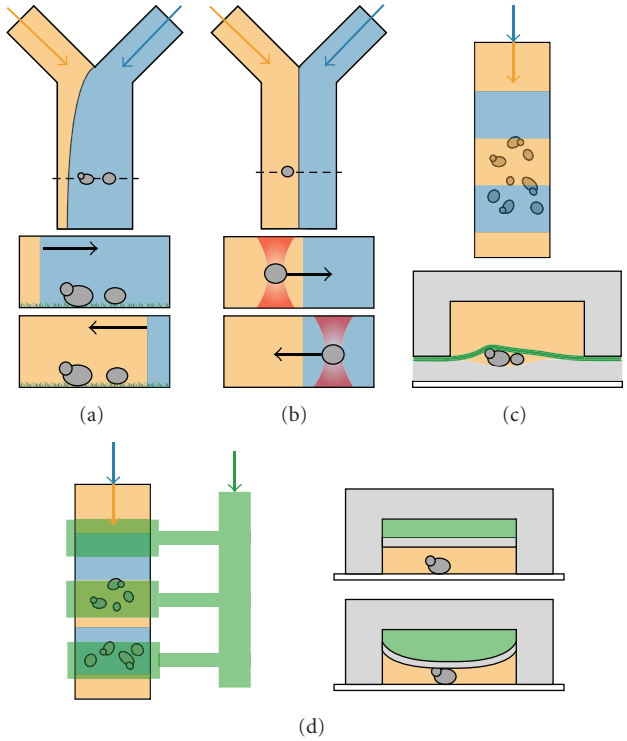


FIGURE 4: Different microfluidics techniques to control the chemical environment of single yeast cells while imaging them through microscopy. (a) Microfluidic system as described in Hersen et al. [11]. Yeast cells are fixed in the channel by the lectin protein Concanavalin A. One inlet is filled with an iso-osmotic media (blue) and the other with a hyperosmotic media (orange). By tightly controlling the pressure in each inlet, it is possible to create a periodic shock on the cells. (b) Optical tweezers system (red) as described by Eriksson et al. permits to control the cells position in the channel with two fluids flowing side by side [12]. (c) The system developed by Charvin et al. uses a dialysis membrane (green) to trap cells on top of a soft PDMS slice [13]. (d) Multilayer microfluidic device [14]. The top layer (green) is used to capture cells. By controlling the pressure inside this channel, cells can be optimally trapped while subjected to periodic shocks. The bottom layer is used to culture cells.

capture the dynamical properties of the pathway, such as feedback loops and signal processing abilities, without taking into account all the details of the biochemical reactions. These approaches require designing experimental systems in which the extracellular environment can be quickly and precisely varied. We will now review the innovative methodologies that have been recently used to study single yeast cells in time varying environments. Then, we will review how those measurements have been integrated into minimal modeling to further study the dynamics of the HOG pathway.

6. Fast Control of the Chemical Environment of Single Cells

Several approaches, using microfluidics [8, 9, 84, 85], have been recently proposed to allow for a fast and reliable control of the chemical environment of yeast cells [7]. Hersen et al. [11] designed a fast binary switch to repeatedly change the environment of single yeast cells between two chemical conditions as fast as every second (Figure 4(a)). They used a Y-shaped flow chamber, 50 μm high and 500 μm wide, with two inlets. One inlet was filled with an isotonic medium, and the other with the same culture medium complemented with sorbitol to increase its osmolarity. At such small scales, flows are laminar and fluids do not mix but rather simply flow side by side. The lateral position of the fluids interface is set by the relative hydrostatic pressure—or the relative flux—of the two inlets. Changing this pressure difference displaces the interface laterally in less than a second. Yeast cells, previously fixed in the channel through concanavalin-A coating were then repeatedly switched from an isotonic to a hyperosmotic environment. An interesting alternative developed by Eriksson et al. [12] consists of moving the cells with optical tweezers (Figure 4(b)) rather than moving fluids over fixed cells. This strategy removes the potential influence of cell adhesion on signaling dynamics related to morphological changes, but at the cost of technological complexity. Also, such a strategy is very time consuming. Holographic tweezers—a sophisticated version of optical tweezers—can help to increase the number of cells that can be observed in real time [86]. Another strategy was proposed by Charvin et al. [13, 87]. Yeast cells are fixed between a permeable dialysis membrane and a cover slip coated with a very thin layer of soft PDMS (Poly-Di-MethylSiloxane). A channel is placed on top of the membrane and allows flow of fresh media and exchange within a few minutes. Nutrients and other chemicals can freely diffuse through the membrane. With this device, environmental exchange happens more slowly, but cells can grow over several generations in a monolayer simplifying their observation through microscopy. Indeed, Charvin et al. used it to force periodic expression of cyclins in yeast growing exponentially up to 8–10 generations.

More complex devices have been proposed, though they require a high degree of expertise to fabricate and manipulate. Bennet et al. [88] developed an environmental switcher capable of generating sinusoidal inputs. Their multilayer device was composed of a microchemostat, with a depth of 4 μm to force yeast cells to grow in a monolayer, and a fluid mixer to generate complex time varying environmental signals for the cells in the chemostat chamber. They used this device, in a particularly elegant work, to revisit the wiring of the GAL system in yeast, by subjecting cells to sinusoidal inputs of carbon source over a range of frequencies. Taylor et al. [14] described a high throughput microfluidics single-cell imaging platform to study the dynamics of the pheromone response in yeast. They combined a fluidic multiplexer, an array of channels, and many sieve valves to trap cells and to control fluid delivery. They were able to perform simultaneous time lapse imaging of 256 chambers with 8 different genotypes with several dynamical inputs. Such a strategy, although very sophisticated, can enhance dramatically the quantity of data gathered to improve our knowledge and refine modeling of MAPK pathways in yeast [7].

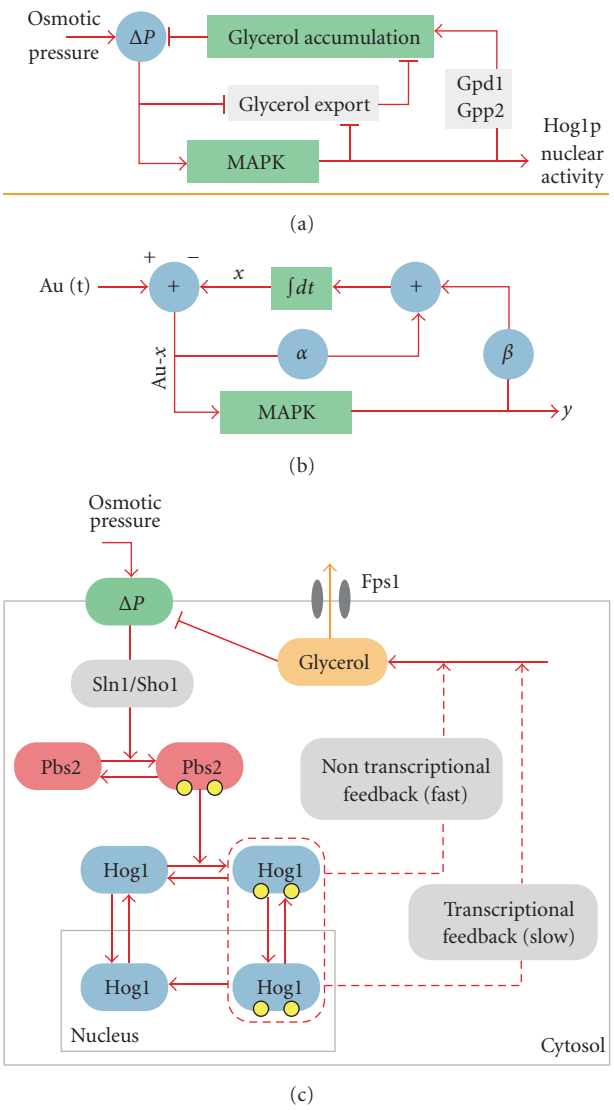


FIGURE 5: Schematic representation of the Hog pathway models of Mettetal et al. [10] and Zi et al. [15]. Pictures are redrawn from original figures of these papers. Top: (a) Diagrammatic representation of Mettetal's model. Au(t) represents the osmolarity applied at time t and the variables x and y can be identified with the intracellular glycerol concentration and the enrichment of Hog1 in the nucleus. The model contains a feedback depending on Hog1p (with strength β) and one, which is independent of Hog1p (strength α). The equations for this model read $\dot{y} = (A_0u - x) - \gamma y$ and $\dot{x} = \alpha(A_0u - x) + \beta y$. (b) The same model, interpreted in biological terms. The export of osmolytes is regulated by a mechanism, which does not depend on the MAPK pathway (e.g., closure of Fps1p) and by a mechanism depending on Hog1p activation. (c) Diagram of the model structure proposed by Zi et al. The model includes a simplified version of the MAPK pathway as well as two different feedbacks induced by activated Hog1p (a slow transcriptional and a fast nontranscriptional). Both of these feedbacks act by increasing the production of glycerol.

7. New Insights from Coupling Complex Stimulus and Reverse Systems Engineering

Using such microfluidics strategies (Figure 4(a)), Hersen et al. studied the HOG pathway response to periodical osmotic stimulation over a range of frequencies. Interestingly, the HOG pathway acts as a low-pass filter, meaning that the output of the pathway (Hog1p nuclear localization) does not follow a fast varying input precisely, but rather integrates fast fluctuations over time. For wild-type strains, when the input signal varies slower than once every 200 s, Hog1p cytoplasmic—nuclear shuttling follows the input variations faithfully [11, 17]. However, when the input varies more rapidly than every 200 s, Hog1p nuclear translocation no longer follows the input faithfully, but instead integrates over the input fluctuations [11, 17]. This typical time is also the slowest time (or limiting step) of activation of the pathway although it was not possible from these experiments to point out which biochemical step was limiting. By genetic removal of one of the two branches, the contribution of each branch was also measured by Hersen et al., and it was found that the SHO1 branch is slower than the SLN1 branch by

almost a factor two. The SHO1 branch was actually unable to integrate the too fast variations of the input whereas the SLN1 branch, when taken alone, was displaying a similar behavior than wild-type cells [11]. Those investigations clearly evidenced that the pathway can be turned off very quickly and repeatedly, suggesting the existence of several feedback loops acting on different timescales.

An attempt to decipher the dynamical aspects of these feedback loops has been done by Mettetal et al. [10], who also examined the response of the Hog1p nuclear localization in response to an oscillating input. They constructed, based on these frequency experiments, a simple predictive model, which was not based on biological knowledge (Figure 5(a)). Subsequently, they identified the two variables of their model with the intercellular osmolyte concentration and the phosphorylation state of Hog1p and concluded that the pathway contains a Hog1-dependent and a Hog1-independent feedback mechanism. By underexpressing Pbs2p, thereby reducing the sensitivity of the Hog1-response to the input, they were able to isolate the Hog1-independent feedback from the Hog1-dependent feedback. Based on this they concluded that the Hog1-dependent feedback is required for fast pathway inactivation. By inhibiting translation, they showed indeed that the slow transcriptional response triggered by Hog1p is only necessary for the adaptation to multiple osmotic shocks, while for a single osmotic shock faster nontranscriptional feedback mechanisms dominate the response. Their conclusion is in perfect agreement with recent experimental investigations showing that even cells with Hog1p anchored to the membrane present an increase of glycerol production after a hyperosmotic shock [89]. Although the details are not known, Hog1p directly or indirectly activates the 6-Phosphofructo-2-kinase (*PFK2*) [68] which leads to an increase production of glycerol through Gpd1p activity.

Hao et al. also focused on rapid non-transcriptional feedback loops. First, they noticed that the response of the SHO1 branch is more transient than that of the SLN1 branch. Then, based on previous observations, they constructed three simple mathematical models, each describing another possible mechanism of HOG inactivation. One model was based on Hog1p mediating activation of a negative regulator (phosphatases), while the other two models focused on the negative control of a positive regulator. Analysis of the different models suggested a Hog1p-dependent feedback mechanism occurring early in the response. Their experimental analysis confirmed this and suggested that Hog1p acts negatively on Sho1p by phosphorylation, thereby implementing a direct negative feedback loop.

Muzzey et al. [18] followed a similar approach to study the feedback mechanisms within the pathway. They identified the transient activation of Hog1p with a feature called perfect adaptation, which states that the steady state output of the pathway does not depend on the strength of the osmotic shock. They argued that robust perfect adaptation requires at least one negative feedback loop containing an integrating component [90] and they analyzed the location of this integrator. They defined an integrating component as a dynamic variable whose rate of change does not depend on itself. They monitored multiple system quantities (cell volume, Hog1p, and glycerol) and used varied input waveforms to analyze the pathway. Similar to Hao et al. [16], they constructed different variants of a mathematical model, each with a different location of the integrating component. The authors found that the integral feedback property is Hog1p dependent and regulates glycerol uptake.

More recently, Zi et al. [15] analyzed the experimental frequency response of the HOG pathway done by Hersen et al. and Mettetal et al. They constructed a minimal model that can reproduce the response of the pathway to oscillating inputs (Figure 5(b)) [15]. They defined a signal response gain, which is defined as the ratio of the integrated change of the output of the pathway to the integrated input change and represents a measurement for the efficiency of signal transduction. They concluded that yeast cells have optimized this signal response gain with respect to certain durations and frequencies of osmotic variations.

These different analyses have shown that the HOG signaling cascade can be described in a very simple and modular way with several feedback loops operating to deactivate the pathway: two operating on short time scales through Hog1p activity (Sho1p deactivation and glycerol production increase), and one depending on transcriptional activation of *GPD1*. The dynamics of the pathway was also precisely measured and it was shown that it behaves as a low-pass filter with a cutoff frequency, probably set by protein concentration. Interestingly, the SHO1 branch which is known to be involved in other cellular processes was shown to be slower in activating the Hog1p MAPK than the SLN1 branch. Finally, those approaches have provided us with an easily tractable mathematical model of the HOG pathway that can be efficiently coupled to detailed mechanistic models to study *in silico* the behavior of this MAPK pathway. Taken together, the coupling between mathematical modeling and experimental frequency analysis of the HOG pathway has given very important insights into the HOG pathway dynamics and more generally its functioning, demonstrating the interest of developing such strategies for studying signaling pathways in yeast.

8. Future Directions

Although the structure and the dynamics of the HOG signaling pathway are now well understood, several key points remain to be elucidated, the most elusive one being the mechanistic functioning of the two osmosensors, Sln1p and the Sho1p complex. Another important aspect of a better understanding of the HOG pathway is to integrate its behavior with other cellular processes. In particular, in 2000, Gasch et al. [73] compiled genome expression profiles of *S. cerevisiae* yeast subjected to several stress conditions and discovered that genes normally induced after a hyperosmotic shock are downregulated in response to a hypo-osmotic shock and vice versa. The CWI pathway is activated by hypo-osmotic stimulation [29], its physiological role being to reinforce the cell wall and prevent the cell from bursting. HOG and CWI do not share direct components but were

seen to interact with each other [91, 92]. During cell growth both pathways may well be activated and deactivated within short intervals to balance between cell expansion and cell wall development. The Sln1p-dependent response regulator Skn7p [93, 94] could have a role in linking the cell-integrity pathway to the HOG pathway. Skn7p also interacts with Rho1p an upstream component of the CWI pathway. The evidence that Skn7p is apparently controlled by sensors of both the HOG pathway and the cell-integrity pathway makes Skn7p an excellent candidate for a regulator that coordinates osmoregulation and cell wall biogenesis [23, 93, 94]. More work is needed to better understand the putative role of Skn7p in coordinating different aspects of turgor pressure control and cell surface assembly. Using minimal models and fluctuating environments to activate periodically the CWI and/or the HOG pathway is one interesting way to explore their interactions. Similarly, it is known that the HOG pathway and the Pheromone pathway can interact [95–98]. For example, a *hog1*Δ strain will respond to a hyperosmotic shock by activating the response to pheromone pathway. Again, the dynamics of such cross-talk has not been intensely studied. Performing time varying inputs with both pheromone and hyperosmotic medium will provide invaluable experimental data to probe for the dynamical aspects of cross-talk between MAPK in yeast.

Since MAPKs pathways are highly conserved from yeast to mammalian cells, it would be interesting to test higher eukaryotic cells, in single cell experiments, for similar system level properties. Although more difficult to implement than for yeast cells, microfluidic technics can also be used to control the external environments of mammalian cells both in time and space. Transposing the approaches described here to mammalian cells will probably give further insights in their signaling pathways dynamics.

9. Conclusion

Since its initial discovery in 1993 [24], extensive molecular and genetic research has uncovered the molecular actors, interactions, and functions of the components in the HOG signaling pathway. However, these methods are limited in that one cannot predict the behavior of a complex system from the analysis of isolated components. Understanding of the entire system requires the use of novel techniques borrowed from engineering, physics, and mathematics. Microfluidic technologies combined with live-cell microscopy have allowed the use of temporally complex stimuli to interrogate pathway function. Kinetic information obtained through biochemistry combined with knowledge of the molecular components has allowed for complex quantitative models of the HOG pathway to be constructed. These models in turn provide experimentally testable predictions about pathway behavior and function. Simple "black-box" models designed to mimic only key components of the pathway have proven useful for understanding specific phenomena. Thus, genetic and biochemical data combined with novel experimental approaches and modeling have allowed for the prediction of the dynamics and systems-level properties of HOG pathway signaling processes. These techniques are easily extended to other signaling pathways of interests with the final goal being to understand the relationships between structure, kinetics, and dynamics at the systems-level in complex biological networks.

Acknowledgment

P. Hersen is supported by the ANR program of the French Government (ANR-JCh-DiSiP). A. Miermont and J. Uhlendorf are students of the Frontier in Life Sciences PhD program (Paris, France).

References

[1] L. H. Hartwell, J. J. Hopfield, S. Leibler, and A. W. Murray, "From molecular to modular cell biology," *Nature*, vol. 402, no. 6761, pp. C47–C52, 1999.

[2] P. Nurse, "Life, logic and information," *Nature*, vol. 454, no. 7203, pp. 424–426, 2008.

[3] R. Brent, "Cell signaling: what is the signal and what information does it carry?" *FEBS Letters*, vol. 583, no. 24, pp. 4019–4024, 2009.

[4] N. L. Novère, M. Hucka, H. Mi et al., "The systems biology graphical notation," *Nature Biotechnology*, vol. 27, no. 8, pp. 735–741, 2009.

[5] A. Jovic, B. Howell, and S. Takayama, "Timing is everything: using fluidics to understand the role of temporal dynamics in cellular systems," *Microfluidics and Nanofluidics*, vol. 6, no. 6, pp. 717–729, 2009.

[6] O. Lipan and W. H. Wong, "The use of oscillatory signals in the study of genetic networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 20, pp. 7063–7068, 2005.

[7] S. Paliwal, J. Wang, and A. Levchenko, "Pulsing cells: how fast is too fast?" *HFSP Journal*, vol. 2, no. 5, pp. 251–256, 2008.

[8] G. M. Whitesides, E. Ostuni, S. Takayama, X. Jiang, and D. E. Ingber, "Soft lithography in biology and biochemistry," *Annual Review of Biomedical Engineering*, vol. 3, pp. 335–373, 2001.

[9] M. R. Bennett and J. Hasty, "Microfluidic devices for measuring gene network dynamics in single cells," *Nature Reviews Genetics*, vol. 10, no. 9, pp. 628–638, 2009.

[10] J. T. Mettetal, D. Muzzey, C. Gómez-Uribe, and A. van Oudenaarden, "The frequency dependence of osmo-adaptation in *Saccharomyces cerevisiae*," *Science*, vol. 319, no. 5862, pp. 482–484, 2008.

[11] P. Hersen, M. N. McClean, L. Mahadevan, and S. Ramanathan, "Signal processing by the HOG MAP kinase pathway," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 20, pp. 7165–7170, 2008.

[12] E. Eriksson, J. Enger, B. Nordlander et al., "A microfluidic system in combination with optical tweezers for analyzing rapid and reversible cytological alterations in single cells upon environmental changes," *Lab on a Chip*, vol. 7, no. 1, pp. 71–76, 2007.

[13] G. Charvin, F. R. Cross, and E. D. Siggia, "A microfluidic device for temporally controlled gene expression and long-term fluorescent imaging in unperturbed dividing yeast cells," *PLoS One*, vol. 3, no. 1, Article ID e1468, 2008.

[14] R. J. Taylor, D. Falconnet, A. Niemistö et al., "Dynamic analysis of MAPK signaling using a high-throughput microfluidic

single-cell imaging platform," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 10, pp. 3758–3763, 2009.

[15] Z. Zi, W. Liebermeister, and E. Klipp, "A quantitative study of the Hog1 MAPK Response to fluctuating osmotic stress in *Saccharomyces cerevisiae*," *PLoS One*, vol. 5, no. 3, Article ID e9522, 2010.

[16] N. Hao, M. Behar, S. C. Parnell et al., "A systems-biology analysis of feedback inhibition in the Sho1 osmotic-stress-response pathway," *Current Biology*, vol. 17, no. 8, pp. 659–667, 2007.

[17] M. N. McClean, P. Hersen, and S. Ramanathan, "In vivo measurement of signaling cascade dynamics," *Cell Cycle*, vol. 8, no. 3, pp. 373–376, 2009.

[18] D. Muzzey, C. A. Gómez-Uribe, J. T. Mettetal, and A. van Oudenaarden, "A systems-level analysis of perfect adaptation in yeast osmoregulation," *Cell*, vol. 138, no. 1, pp. 160–171, 2009.

[19] C. Widmann, S. Gibson, M. B. Jarpe, and G. L. Johnson, "Mitogen-activated protein kinase: conservation of a three-kinase module from yeast to human," *Physiological Reviews*, vol. 79, no. 1, pp. 143–180, 1999.

[20] D. Sheikh-Hamad and M. C. Gustin, "MAP kinases and the adaptive response to hypertonicity: functional preservation from yeast to mammals," *American Journal of Physiology*, vol. 287, no. 6, pp. F1102–F1110, 2004.

[21] R. E. Chen and J. Thorner, "Function and regulation in MAPK signaling pathways: lessons learned from the yeast *Saccharomyces cerevisiae*," *Biochimica et Biophysica Acta*, vol. 1773, no. 8, pp. 1311–1340, 2007.

[22] M. C. Gustin, J. Albertyn, M. Alexander, and K. Davenport, "Map kinase pathways in the yeast *Saccharomyces cerevisiae*," *Microbiology and Molecular Biology Reviews*, vol. 62, no. 4, pp. 1264–1300, 1998.

[23] S. Hohmann, "Osmotic stress signaling and osmoadaptation in yeasts," *Microbiology and Molecular Biology Reviews*, vol. 66, no. 2, pp. 300–372, 2002.

[24] J. L. Brewster, T. De Valoir, N. D. Dwyer, E. Winter, and M. C. Gustin, "An osmosensing signal transduction pathway in yeast," *Science*, vol. 259, no. 5102, pp. 1760–1763, 1993.

[25] S. M. O'Rourke, I. Herskowitz, and E. K. O'Shea, "Yeast go the whole HOG for the hyperosmotic response," *Trends in Genetics*, vol. 18, no. 8, pp. 405–412, 2002.

[26] P. J. Westfall, D. R. Ballon, and J. Thorner, "When the stress of your environment makes you go HOG wild," *Science*, vol. 306, no. 5701, pp. 1511–1512, 2004.

[27] S. Hohmann, "Control of high osmolarity signalling in the yeast *Saccharomyces cerevisiae*," *FEBS Letters*, vol. 583, no. 24, pp. 4025–4029, 2009.

[28] S. Hohmann, M. Krantz, and B. Nordlander, "Yeast osmoregulation," *Methods in Enzymology*, vol. 428, pp. 29–45, 2007.

[29] K. R. Davenport, M. Sohaskey, Y. Kamada, D. E. Levin, and M. C. Gustin, "A second osmosensing signal transduction pathway in yeast: hypotonic shock activates the PKC1 protein kinase-regulated cell integrity pathway," *Journal of Biological Chemistry*, vol. 270, no. 50, pp. 30157–30161, 1995.

[30] H. Martín, J. M. Rodríguez-Pachón, C. Ruiz, C. Nombela, and M. Molina, "Regulatory mechanisms for modulation of signaling through the cell integrity Slt2-mediated pathway in *Saccharomyces cerevisiae*," *Journal of Biological Chemistry*, vol. 275, no. 2, pp. 1511–1519, 2000.

[31] J. C. Harrison, E. S. G. Bardes, Y. Ohya, and D. J. Lew, "A role for the Pkc1p/Mpk1p kinase cascade in the morphogenesis checkpoint," *Nature Cell Biology*, vol. 3, no. 4, pp. 417–420, 2001.

[32] E. A. Elion, "Pheromone response, mating and cell biology," *Current Opinion in Microbiology*, vol. 3, no. 6, pp. 573–581, 2000.

[33] K. Tedford, S. Kim, D. Sa, K. Stevens, and M. Tyers, "Regulation of the mating pheromone and invasive growth responses in yeast by two MAP kinase substrates," *Current Biology*, vol. 7, no. 4, pp. 228–238, 1997.

[34] F. Banuett, "Signaling in the yeasts: an informational casacade with links to the filamentous fungi," *Microbiology and Molecular Biology Reviews*, vol. 62, no. 2, pp. 249–274, 1998.

[35] B. Ren, F. Robert, J. J. Wyrick et al., "Genome-wide location and function of DNA binding proteins," *Science*, vol. 290, no. 5500, pp. 2306–2309, 2000.

[36] V. Gavrias, A. Andrianopoulos, C. J. Gimeno, and W. E. Timberlake, "*Saccharomyces cerevisiae* TEC1 is required for pseudohyphal growth," *Molecular Microbiology*, vol. 19, no. 6, pp. 1255–1263, 1996.

[37] M. Proft, A. Pascual-Ahuir, E. de Nadal, J. Arño, R. Serrano, and F. Posas, "Regulation of the Sko1 transcriptional repressor by the Hog1 MAP kinase in response to osmotic stress," *The EMBO Journal*, vol. 20, no. 5, pp. 1123–1133, 2001.

[38] P. M. Alepuz, E. de Nadal, M. Zapater, G. Ammerer, and F. Posas, "Osmostress-induced transcription by Hot1 depends on a Hog1-mediated recruitment of the RNA Pol II," *The EMBO Journal*, vol. 22, no. 10, pp. 2433–2442, 2003.

[39] P. Shore and A. D. Sharrocks, "The MADS-box family of transcription factors," *European Journal of Biochemistry*, vol. 229, no. 1, pp. 1–13, 1995.

[40] K. Madden, Y. J. Sheu, K. Baetz, B. Andrews, and M. Snyder, "SBF cell cycle regulator as a target of the yeast PKC-MAP kinase pathway," *Science*, vol. 275, no. 5307, pp. 1781–1784, 1997.

[41] K. Doi, A. Gartner, G. Ammerer et al., "MSG5, a novel protein phosphatase promotes adaptation to pheromone response in S.cerevisiae," *The EMBO Journal*, vol. 13, no. 1, pp. 61–70, 1994.

[42] X. L. Zhan, R. J. Deschenes, and K. L. Guan, "Differential regulation of FUS3 map kinase by tyrosine-specific phosphatases PTP2/PTP3 and dual-specificity phosphatase MSG5 in *Saccharomyces cerevisiae*," *Genes & Development*, vol. 11, no. 13, pp. 1690–1702, 1997.

[43] C. P. Mattison and I. M. Ota, "Two protein tyrosine phosphatases, Ptp2 and Ptp3, modulate the subcellular localization of the Hog1 MAP kinase in yeast," *Genes & Development*, vol. 14, no. 10, pp. 1229–1235, 2000.

[44] J. Warmka, J. Hanneman, J. Lee, D. Amin, and I. Ota, "Ptc1, a type 2C Ser/Thr phosphatase, inactivates the HOG pathway by dephosphorylating the mitogen-activated protein kinase Hog1," *Molecular and Cellular Biology*, vol. 21, no. 1, pp. 51–60, 2001.

[45] T. Maeda, A. Y. M. Tsai, and H. Saito, "Mutations in a protein tyrosine phosphatase gene (PTP2) and a protein serine/threonine phosphatase gene (PTC1) cause a synthetic growth defect in *Saccharomyces cerevisiae*," *Molecular and Cellular Biology*, vol. 13, no. 9, pp. 5408–5417, 1993.

[46] H. Martín, M. Flández, C. Nombela, and M. Molina, "Protein phosphatases in MAPK signalling: we keep learning from yeast," *Molecular Microbiology*, vol. 58, no. 1, pp. 6–16, 2005.

[47] E. de Nadal, P. M. Alepuz, and F. Posas, "Dealing with osmostress through MAP kinase activation," *EMBO Reports*, vol. 3, no. 8, pp. 735–740, 2002.

[48] P. D'Haeseleer, "Closing the circle of osmoregulation," *Nature Biotechnology*, vol. 23, no. 8, pp. 941–942, 2005.

[49] H. Saito and K. Tatebayashi, "Regulation of the osmoregulatory HOG MAPK cascade in yeast," *Journal of Biochemistry*, vol. 136, no. 3, pp. 267–272, 2004.

[50] T. Maeda, S. M. Wurgler-Murphy, and H. Saito, "A two-component system that regulates an osmosensing MAP kinase cascade in yeast," *Nature*, vol. 369, no. 6477, pp. 242–245, 1994.

[51] F. Posas, S. M. Wurgler-Murphy, T. Maeda, E. A. Witten, T. C. Thai, and H. Saito, "Yeast HOG1 MAP kinase cascade is regulated by a multistep phosphorelay mechanism in the SLN1-YPD1-SSK1 "two-component" osmosensor," *Cell*, vol. 86, no. 6, pp. 865–875, 1996.

[52] V. Reiser, D. C. Raitt, and H. Saito, "Yeast osmosensor Sln1 and plant cytokinin receptor Cre1 respond to changes in turgor pressure," *Journal of Cell Biology*, vol. 161, no. 6, pp. 1035–1040, 2003.

[53] J. Macia, S. Regot, T. Peeters, N. Conde, R. Solé, and F. Posas, "Dynamic signaling in the Hog1 MAPK pathway relies on high basal signal transduction," *Science Signaling*, vol. 2, no. 63, p. ra13, 2009.

[54] T. Maeda, M. Takekawa, and H. Saito, "Activation of yeast PBS2 MAPKK by MAPKKKs or by binding of an SH3-containing osmosensor," *Science*, vol. 269, no. 5223, pp. 554–558, 1995.

[55] V. Reiser, S. M. Salah, and G. Ammerer, "Polarized localization of yeast Pbs2 depends on osmostress, the membrane protein Sho1 and Cdc42," *Nature Cell Biology*, vol. 2, no. 9, pp. 620–627, 2000.

[56] B. T. Seet and T. Pawson, "MAPK signaling: Sho business," *Current Biology*, vol. 14, no. 17, pp. R708–R710, 2004.

[57] E. de Nadal, F. X. Real, and F. Posas, "Mucins, osmosensors in eukaryotic cells?" *Trends in Cell Biology*, vol. 17, no. 12, pp. 571–574, 2007.

[58] K. Tatebayashi, K. Tanaka, H. Y. Yang et al., "Transmembrane mucins Hkr1 and Msb2 are putative osmosensors in the SHO1 branch of yeast HOG pathway," *The EMBO Journal*, vol. 26, no. 15, pp. 3521–3533, 2007.

[59] S. M. O'Rourke and I. Herskowitz, "A third osmosensing branch in *Saccharomyces cerevisiae* requires the Msb2 protein and functions in parallel with the Sho1 branch," *Molecular and Cellular Biology*, vol. 22, no. 13, pp. 4739–4749, 2002.

[60] T. Yabe, T. Yamada-Okabe, S. Kasahara et al., "HKR1 encodes a cell surface protein that regulates both cell wall β-glucan synthesis and budding pattern in the yeast *Saccharomyces cerevisiae*," *Journal of Bacteriology*, vol. 178, no. 2, pp. 477–483, 1996.

[61] M. Hayashi and T. Maeda, "Activation of the HOG pathway upon cold stress in *Saccharomyces cerevisiae*," *Journal of Biochemistry*, vol. 139, no. 4, pp. 797–803, 2006.

[62] S. J. Mansour, K. A. Resing, J. M. Candi et al., "Mitogen-activated protein (MAP) kinase phosphorylation of MAP kinase kinase: determination of phosphorylation sites by mass spectrometry and site-directed mutagenesis," *Journal of Biochemistry*, vol. 116, no. 2, pp. 304–314, 1994.

[63] M. Krantz, D. Ahmadpour, L. G. Ottosson et al., "Robustness and fragility in the yeast high osmolarity glycerol (HOG) signal-transduction pathway," *Molecular Systems Biology*, vol. 5, article no. 281, 2009.

[64] A. Zarrinpar, R. P. Bhattacharyya, M. P. Nittler, and W. A. Lim, "Sho1 and Pbs2 act as coscaffolds linking components in the yeast high osmolarity MAP kinase pathway," *Molecular Cell*, vol. 14, no. 6, pp. 825–832, 2004.

[65] P. Ferrigno, F. Posas, D. Koepp, H. Saito, and P. A. Silver, "Regulated nucleo/cytoplasmic exchange of HOG1 MAPK requires the importin β homologs NMD5 and XPO1," *The EMBO Journal*, vol. 17, no. 19, pp. 5606–5614, 1998.

[66] M. Mollapour and P. W. Piper, "Hog1 mitogen-activated protein kinase phosphorylation targets the yeast Fps1 aquaglyceroporin for endocytosis, thereby rendering cells resistant to acetic acid," *Molecular and Cellular Biology*, vol. 27, no. 18, pp. 6446–6456, 2007.

[67] E. Bilsland-Marchesan, J. Ariño, H. Saito, P. Sunnerhagen, and F. Posas, "Rck2 kinase is a substrate for the osmotic stress-activated mitogen-activated protein kinase Hog1," *Molecular and Cellular Biology*, vol. 20, no. 11, pp. 3887–3895, 2000.

[68] H. Dihazi, R. Kessler, and K. Eschrich, "High osmolarity glycerol (HOG) pathway-induced phosphorylation and activation of 6-phosphofructo-2-kinase are essential for glycerol accumulation and yeast cell proliferation under hyperosmotic stress," *Journal of Biological Chemistry*, vol. 279, no. 23, pp. 23961–23968, 2004.

[69] M. J. Tamás, K. Luyten, F. C. W. Sutherland et al., "Fps1p controls the accumulation and release of the compatible solute glycerol in yeast osmoregulation," *Molecular Microbiology*, vol. 31, no. 4, pp. 1087–1104, 1999.

[70] M. Siderius, O. Van Wuytswinkel, K. A. Reijenga, M. Kelders, and W. H. Mager, "The control of intracellular glycerol in *Saccharomyces cerevisiae* influences osmotic stress response and resistance to increased temperature," *Molecular Microbiology*, vol. 36, no. 6, pp. 1381–1390, 2000.

[71] S. E. Beese, T. Negishi, and D. E. Levin, "Identification of positive regulators of the yeast Fps1 glycerol channel," *PLoS Genetics*, vol. 5, no. 11, Article ID e1000738, 2009.

[72] J. Yale and H. J. Bohnert, "Transcript expression in *Saccharomyces cerevisiae* at high salinity," *Journal of Biological Chemistry*, vol. 276, no. 19, pp. 15996–16007, 2001.

[73] A. P. Gasch, P. T. Spellman, C. M. Kao et al., "Genomic expression programs in the response of yeast cells to environmental changes," *Molecular Biology of the Cell*, vol. 11, no. 12, pp. 4241–4257, 2000.

[74] M. Rep, M. Krantz, J. M. Thevelein, and S. Hohmann, "The transcriptional response of *Saccharomyces cerevisiae* to osmotic shock. Hot1p and Msn2p/Msn4p are required for the induction of subsets of high osmolarity glycerol pathway-dependent genes," *Journal of Biological Chemistry*, vol. 275, no. 12, pp. 8290–8300, 2000.

[75] P. M. Alepuz, A. Jovanovic, V. Reiser, and G. Ammerer, "Stress-induced MAP kinase Hog1 is part of transcription activation complexes," *Molecular Cell*, vol. 7, no. 4, pp. 767–777, 2001.

[76] E. de Nadal, L. Casadomé, and F. Posas, "Targeting the MEF2-like transcription factor Smp1 by the stress-activated Hog1 mitogen-activated protein kinase," *Molecular and Cellular Biology*, vol. 23, no. 1, pp. 229–237, 2003.

[77] M. Rep, M. Proft, F. Remize et al., "The *Saccharomyces cerevisiae* Sko1p transcription factor mediates HOG pathway-dependent osmotic regulation of a set of genes encoding enzymes implicated in protection from oxidative damage," *Molecular Microbiology*, vol. 40, no. 5, pp. 1067–1083, 2001.

[78] A. Valadi, K. Granath, L. Gustafsson, and L. Adler, "Distinct intracellular localization of Gpd1p and Gpd2p, the two yeast isoforms of NAD-dependent glycerol-3-phosphate dehydrogenase, explains their different contributions to redox-driven glycerol production," *Journal of Biological Chemistry*, vol. 279, no. 38, pp. 39677–39685, 2004.

[79] J. Albertyn, S. Hohmann, J. M. Thevelein, and B. A. Prior, "GPD1, which encodes glycerol-3-phosphate dehydrogenase,

is essential for growth under osmotic stress in *Saccharomyces cerevisiae*, and its expression is regulated by the high-osmolarity glycerol response pathway," *Molecular and Cellular Biology*, vol. 14, no. 6, pp. 4135–4144, 1994.

[80] C. Ferreira, F. Van Voorst, A. Martins et al., "A member of the sugar transporter family, Stl1p is the glycerol/H symporter in *Saccharomyces cerevisiae*," *Molecular Biology of the Cell*, vol. 16, no. 4, pp. 2068–2076, 2005.

[81] E. Klipp, B. Nordlander, R. Krüger, P. Gennemark, and S. Hohmann, "Integrative model of the response of yeast to osmotic shock," *Nature Biotechnology*, vol. 23, no. 8, pp. 975–982, 2005.

[82] P. Gennemark, B. Nordlander, S. Hohmann, and D. Wedelin, "A simple mathematical model of adaptation to high osmolarity in yeast," *In Silico Biology*, vol. 6, no. 3, pp. 193–214, 2006.

[83] S. Karlgren, N. Pettersson, B. Nordlander et al., "Conditional osmotic stress in yeast: a system to study transport through aquaglyceroporins and osmostress signaling," *Journal of Biological Chemistry*, vol. 280, no. 8, pp. 7186–7193, 2005.

[84] P. Wilding, J. Pfahler, H. H. Bau, J. N. Zemel, and L. J. Kricka, "Manipulation and flow of biological fluids in straight channels micromachined in silicon," *Clinical Chemistry*, vol. 40, no. 1, pp. 43–47, 1994.

[85] Y. Xia and G. M. Whitesides, "Soft lithography," *Annual Review of Materials Science*, vol. 28, no. 1, pp. 153–184, 1998.

[86] A. Lafong, W. J. Hossack, J. Arlt, T. J. Nowakowski, and N. D. Read, "Time-Multiplexed Laguerre-Gaussian holographic optical tweezers for biological applications," *Optics Express*, vol. 14, no. 7, pp. 3065–3072, 2006.

[87] G. Charvin, F. R. Cross, and E. D. Siggia, "Forced periodic expression of G cyclins phase-locks the budding yeast cell cycle," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 16, pp. 6632–6637, 2009.

[88] M. R. Bennett, W. L. Pang, N. A. Ostroff et al., "Metabolic gene regulation in a dynamically changing environment," *Nature*, vol. 454, no. 7208, pp. 1119–1122, 2008.

[89] P. J. Westfall, J. C. Patterson, R. E. Chen, and J. Thorner, "Stress resistance and signal fidelity independent of nuclear MAPK function," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 34, pp. 12212–12217, 2008.

[90] T. M. Yi, Y. Huang, M. I. Simon, and J. Doyle, "Robust perfect adaptation in bacterial chemotaxis through integral feedback control," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 9, pp. 4649–4653, 2000.

[91] R. Alonso-Monge, E. Real, I. Wojda, J. P. Bebelman, W. H. Mager, and M. Siderius, "Hyperosmotic stress response and regulation of cell wall integrity in *Saccharomyces cerevisiae* share common functional aspects," *Molecular Microbiology*, vol. 41, no. 3, pp. 717–730, 2001.

[92] R. García, J. M. Rodríguez-Peña, C. Bermejo, C. Nombela, and J. Arroyo, "The high osmotic response and cell wall integrity pathways cooperate to regulate transcriptional responses to zymolyase-induced cell wall stress in *Saccharomyces cerevisiae*," *Journal of Biological Chemistry*, vol. 284, no. 16, pp. 10901–10911, 2009.

[93] J. L. Brown, H. Bussey, and R. C. Stewart, "Yeast Skn7p functions in a eukaryotic two-component regulatory pathway," *The EMBO Journal*, vol. 13, no. 21, pp. 5186–5194, 1994.

[94] J. L. Brown, S. North, and H. Bussey, "SKN7, a yeast multicopy suppressor of a mutation affecting cell wall β-glucan assembly, encodes a product with domains homologous to prokaryotic two-component regulators and to heat shock transcription factors," *Journal of Bacteriology*, vol. 175, no. 21, pp. 6908–6915, 1993.

[95] A. Mody, J. Weiner, and S. Ramanathan, "Modularity of MAP kinases allows deformation of their signalling pathways," *Nature Cell Biology*, vol. 11, no. 4, pp. 484–491, 2009.

[96] L. Rensing and P. Ruoff, "How can yeast cells decide between three activated MAP kinase pathways? A model approach," *Journal of Theoretical Biology*, vol. 257, no. 4, pp. 578–587, 2009.

[97] M. N. McClean, A. Mody, J. R. Broach, and S. Ramanathan, "Cross-talk and decision making in MAP kinase pathways," *Nature Genetics*, vol. 39, no. 3, pp. 409–414, 2007.

[98] S. M. O'Rourke and I. Herskowitz, "The Hog1 MAPK prevents cross talk between the HOG and pheromone response MAPK pathways in *Saccharomyces cerevisiae*," *Genes & Development*, vol. 12, no. 18, pp. 2874–2886, 1998.

Colophon

This thesis was typeset using LaTeX, originally developed by Leslie Lamport and based on Donald Knuth's TeX. The body text is set in 11 point Arno Pro, designed by Robert Slimbach in the style of book types from the Aldine Press in Venice, and issued by Adobe in 2007. A template, which can be used to format a PhD thesis with this look and feel, has been released under the permissive MIT (X11) license, and can be found online at github.com/suchow/ or from the author at suchow@post.harvard.edu.